Entry limiting agreements: first mover advantage, authorized generics and pay-for-delay deals †

Farasat A.S. Bokhari [a,b,*] Franco Mariuzzo [a,b] Arnold Polanski [a]

[a] *School of Economics, University of East Anglia, Norwich NR4 7TJ, UK*

[b] *Centre for Competition Policy, University of East Anglia, Norwich NR4 7TJ, UK*

January 2020

Abstract

During patent litigation, pay-for-delay deals involve a payment from a patent holder of a branded drug to a generic drug manufacturer to delay entry and withdraw the patent challenge. In return for staying out of the market, the generic firm receives a payment, and/or an authorized licensed entry at a later date, but before the patent expiration. We examine why such deals are stable when there are multiple potential entrants. We combine the first mover advantage for the first generic with the ability of the branded manufacturer to launch an authorized generic to show when pay-for-delay deals are an equilibrium outcome. We further show that limiting a branded firm's ability to launch an authorized generic prior to entry by a successful challenger will deter such deals. However, removing exclusivity period for the first generic challenger will not.

Key words: pharmaceuticals, pay-for-delay, reverse payments, authorized generics, first mover advantage

JEL Classification: L41, K21, K41

†This draft has benefitted from comments from Ashish Arora, Gregory Asciolla, Steve Davies, Ginger Zhe Jin, Kai-Uwe Kühn, Margaret Kyle, Emiliya Lazarova, Abhijit Ramalingam, Pierre Régibeau, Fiona Scott-Morton, and Mike Walker. We are also very grateful for useful feedback given by participants at the "Competition Issues in Pharmaceuticals: The Challenges Ahead" workshop organized by the Centre for Competition Policy (University of East Anglia) as well as participants at EARIE (2015), IIOC (2016) and CRESSE (2018). We would also like to thank the associate editor and two anonymous referees of the journal for helping us produce a much improved version. The first draft was circulated on June 29, 2015 with the title, "Entry limiting agreements for pharmaceuticals: pay-for-delay and authorized generic deals".

*Corresponding Author: Farasat Bokhari (f.bokhari@uea.ac.uk)

1. INTRODUCTION

A pay-for-delay deal is a 'reverse payment' from a patent holder to another drug manufacturer seeking entry for its generic equivalent drug. They arise in out of court settlements because the patent holder has sued the potential entrant for infringement of its intellectual property. The deals are referred to as 'reverse payments', because the payment is from the infringed to the infringer, rather than the other way around. In return for the payment, the generic firm abandons its challenge and agrees to stay out of the market. Moreover, it often also acquires a right from the patent holder to enter at a later date, but before the patent expiration itself as an authorized licensed generic with an exclusive license. The branded firm may additionally agree not to launch an in-house generic during the exclusive license period. The eventual entry by a generic firm takes place at a later date, potentially well after a court may have declared the patent invalid, but also typically before the expiration of patent itself.

Prior literature has relied on institutional details of the American legal system vis-à-vis the market authorization rules and provisions of the Hatch-Waxman Act of 1984, particularly section IV of the Act (called a 'para IV challenge') to provide an explanation of how pay-for-delay (P4D) deals come about (Bulow, 2004, Frank, 2007, Hemphill, 2009, Scott Morton and Kyle, 2011, Mulcahy, 2011, Scott Morton, 2013, Regibeau, 2013). As has been noted in this literature, these deals are typically initiated after the patent protecting the molecule expires, but while other patents associated with the drug, as registered by the US Food and Drug Administration (FDA), remain in force. The first generic company to successfully file for market authorization under section IV of the Act is explicitly rewarded a six month exclusivity period, during which time no other generic firm can market its drug. Such a reward is not available to later challengers even if the first one settles with the patent holder.

There are clear trade-offs in arriving at such a deal. The generic firm can reject the deal and face litigation cost and take its chances in the court. If it wins, it can earn duopoly profits for 180-days, followed by an N-opoly period in which it shares the generic segment of the market with other generic producers. Alternatively, it can avoid the uncertainty and litigation costs and accept

a suitable payment. The challenger gains from settling as long as the payment to stay out at least equals the expected future profit net of litigation costs. However, if the patent is strong, the monopolist may not offer a deal as long as the litigation costs are less than its expected monopoly profits. Conversely, a reverse payment settlement that keeps the challenger out of the market is also profitable for the branded firm as it can maintain its monopoly position (Drake et al., 2015, McGuire et al., 2016) and the payment does not exceed the expected difference between monopoly and duopoly profits minus litigation costs. But the settlement can also expose the monopolist to future challenges by other generic producers as it can signal a weak patent.

We focus on the incentives involved in reaching P4D deals before filing for generic entry i.e., *ex ante* P4D deals when no generic can use its first filer status with the FDA to block entries by other generic challengers.[1] The main goal of this paper is to answer the following key question: if the originator can pay off the generic producer to refrain from challenging its patent, and to stay out of the market for some time, how much do they have to pay, and why do other generic challengers not grab the same opportunity to also get paid off? And if indeed this is possible, then how is the initial deal profitable for the originator if there are many potential challengers?

In the *FTC v. Actavis, Inc.* case argued before the US Supreme Court, the 5-3 majority opinion pointed out that the 180-day exclusivity of the Hatch-Waxman Act is precisely why P4D deals are stable.[2] The majority opinion goes on to state that because the 180-day exclusivity is not available to later challengers (even if the first challenger settles in a P4D case), the low potential reward prevents others from seeking entry. Note however that the first to file for a generic entry enjoys no

[1]Prior to 2003, the 180-day award to the first filer could block entry by other generics in a direct manner: if the first filer delays entry, for say three years due to a P4D settlement, then entry for all other generics could not happen for three and half years. Based on this, many had called for the loop-hole to be closed, and indeed the Medicare Modernization Act of 2003 made amendments which can trigger a forfeiture of the exclusivity period if a successful challenger delays entry. The Act also requires the settling parties to submit the terms to the FTC for antitrust review if it relates to the generic application filed with the FDA. If a generic entry application was not already filed with the FDA, the firms would not need to disclose the terms of settlement to FTC, for instance if patent settlement case were resolved in the Patent Trial and Appeal Board. Thus, while the Act solves one aspect of regulatory problems related to P4D deals, it leaves open the path for parties to settle prior to filing for generic entry with the FDA and avoid antitrust scrutiny by the FTC.

[2]The 180-day exclusivity was cited in combination with a 30-month stay order at FDA in case of a challenge. "Would not a high reverse payment signal to other potential challengers that the patentee lacks confidence in its patent, thereby provoking additional challenges, perhaps too many for the patentee to 'buy off?' Two special features of Hatch-Waxman mean that the answer to this question is 'not necessarily so'." US Supreme Court (2013, p.16).

statutory exclusivity period in the European Union, and yet entry limiting reverse payments take place on both sides of the Atlantic.

In this paper we investigate how P4D deals would arise if the 180-day exclusivity were available to the late filers, as in the ex ante settlements, or for instance if exclusivity was awarded to first successful challenger (hereinafter FSC system) rather than to the first filer (hereinafter FF system) as suggested in Hemphill and Lemley (2011). We use two key features of the pharmaceutical industry to provide an explanation for the stability of pay-for-delay deals. The first is a first mover advantage for a generic firm that is distinct from any exclusivity period or even an incumbency period, and arises due to higher willingness to pay for the first generic relative to other generics. The second is the ability of a branded manufacturer to launch a generic, known as an authorized generic (AG), either itself or via a third party under a licensing agreement that can undercut incentives for independent generic entry.

We propose a simple model with one branded firm with a patent and many potential challengers. The branded firm can threaten the first challenger to launch its own generic (an in-house authorized generic, some times known as a 'pseudogeneric') and deprive the challenger of any first mover advantage in the generic segment. However, in this case it would incur a cost associated with acquiring a speciality to successfully market a generic. Alternatively, if this cost is too high, it can offer to pay off the first challenger to stay out of the market. If the deal is accepted, the branded firm can use the first challenger to ward off entry by any subsequent challengers. It can do so by threatening to launch a generic via the first paid-off challenger *prior* to the second challenger's entry in case a patent litigation outcome is in favor of the latter. If at any stage the branded firm chooses to execute the threat (launch an AG), it takes away the challenger's first mover advantage thereby reducing the latter's incentive to contest entry. However, launching an generic either in-house or via the first challenger, also forces the branded firm to enter into a triopoly rather than engage in a competitive duopoly against the winning challenger, and hence the threat may not always be credible.

We provide conditions under which the threat becomes credible. We show that if the first mover advantage exists and is larger than a threshold, then under an endogenously determined licensing fee for the authorized generic (determined via take-it or leave-it offer), the branded firm is better off in a triopoly with the authorized generic than in a competitive duopoly. This is because if the first generic entrant can capture a significantly large share of the generic market, then both the branded firm and the challenger can agree on a licensing fee that allows the launch. Similar reasoning applies to the case when the branded firm can launch its own in-house generic at zero (or low) cost and cannibalize its branded product, the only difference being that it fully captures any profits associated with the sales of the authorized generic rather than a negotiated licensing fee. In the ensuing triopoly, the branded firm gets to recoup some of the losses relative to its favored monopoly position due the sales of the authorized generic via the licensing fee (or all of generic's profit if it was self launched) and hence it is better off than being in a duopoly. Thus for a large enough first mover advantage, the threat to launch an authorized generic either itself or via the first challenger is credible, and working backwards, second and subsequent potential challengers may optimally choose to stay out of the market if their expected profit is lower than the cost of litigation.

When the first mover advantage is not large, subsequent generic firms may choose to challenge entry, in which case it is necessary for the branded firm to make smaller payments to all subsequent challengers to maintain its monopoly position. With just a few potential challengers, the branded firm can pay off all the challengers and still be better off than facing litigation, which may be true is some therapeutic classes. However, if the number of challengers is large, the net surplus from paying off multiple challengers eventually becomes negative and hence no P4D deals are possible when first mover advantage is small.

To demonstrate all this, we first model equilibrium profits and payments with just three firms (brand and two challengers under the first successful challenger system) and show how P4D deals come about when the branded manufacturer can pay off one or both challengers and/or launch its own authorized generic. We then extend the analysis to the case with many challengers and show

that P4D deals are still possible if the first mover advantage is larger than a threshold. We compare these to the case when exclusivity is restricted to either just the first filer (the current first filer system), or when exclusivity period is removed altogether (a policy proposal). Both of these cases give similar outcomes and show that P4D deals are possible over a much larger range of model parameters and that the first mover advantage does not matter in these cases. Finally, we also consider the case where we limit the ability of the branded firm to launch an authorized generic if a subsequent challenger wins patent litigation and show that P4D deals are then not possible.

2. Related Literature

Sequential P4D deals with potential challengers share the logic developed by Bernheim (1984), but with deterrence investment substituted with P4D deals and licensing an authorized generic (AG). Indeed, the strategy of launching an AG via a P4D deal with a challenger is similar to earlier studies that focus on licensing as a strategy to maintain market leadership and/or deter entry. For instance, Gallini (1984) shows the conditions where the incumbent licenses its production technology to a potential entrant in exchange for terminating research into competing for better technology, while Rockett (1990) and Eswaran (1994) provide models where the incumbent licenses either the weaker competitor or a competitor from outside of the industry, so as to crowd the market and discourage stronger competitors from entering. By contrast, in our paper, the generic with the AG license is the *de facto* strongest competitor to the brand as it enters before other generics and grabs the first mover advantage. Additionally, instead of a license being introduced prior to the potential competitor incurring entry costs, in our paper the license is issued and AG launched only if the next potential entrant has incurred an entry cost (i.e., litigation cost), and is successful.

A significant economic and legal literature builds around theory of harm and focuses on the legality of pay-for-delay deals (Shapiro, 2003a, Lemley and Shapiro, 2005, Farrell and Shapiro, 2008, Gratz, 2012). Under Shapiro's antitrust welfare criteria – that a settlement should leave the consumers at least as well off as the ongoing patent litigation – a payment that exceeds the expected litigation costs of the licensor is sufficient to establish that consumers lose from the settlement (Shapiro, 2003b, Elhauge and Krueger, 2012). In line with this reasoning, several authors have argued that

pay-for-delay settlements should carry a presumption of *per se* anticompetitive behavior (see for instance, Hovenkamp et al., 2003, Bulow, 2004, Leffler and Leffler, 2004, Hemphill, 2009). Others have pointed out that while the theory of harm is useful, it has limitations and cannot be applied directly to the more complex agreements between the parties, or that P4D deals can in fact be pro-competitive in some situations, and hence such deals should not be *per se* illegal (Crane, 2002, Willig and Bigelow, 2004, Dickey et al., 2010, Regibeau, 2013).

For instance, Padilla and Meunier (2015) critique Elhauge and Krueger (2012) and claim that if either the assumption of single potential entrant or of complete information regarding beliefs about the strength of the patent (and hence probability of court outcome) is relaxed, then a per se rule that views reverse payment in excess of litigation cost as anti-competitive cannot be justified. In the context of multiple challengers, Palikot and Pietola (2018) consider externalities that arise from settlement as it may increase the probability of a future litigation (settlement reveals that a patent is weak). They find that litigation will happen for intermediate strength patents, but weak and strong patents are settled via licensing or P4D deals. By contrast, in our model, a P4D deal involves a licensing agreement with the settling party which is in line with typical P4D deals (see Hemphill, 2007). Further, while they require expected duopoly profit payments to all potential entrants, in our case it may be necessary to pay all challengers if the first mover advantage is below a threshold. Finally, Marxen and Montez (2018) adopt a model of vertical quality differentiation to study early entry accommodation (as opposed to P4D deal) and find that under price competition early entry agreements between the incumbent of a patented drug and a generic entrant are always welfare improving, whereas if competition is in quantity, these are welfare enhancing only when they come along with low fixed costs of entry. In either of the two competitive conducts, consumer welfare rises if fixed costs of entry are high. While our focus is not on welfare analysis (which we briefly address in the appendix), our model also allows vertical quality differentiation but we do that in the context of a representative consumer and associate it to the order of entry.

An important ingredient in our model is the advantage of first generic relative to other generics. The first mover advantage for the first generic is in part due to the fact that it enters and serves

the market for a longer period of time compared to other generics, but also because it captures and sustains a much larger share of the generic market over a period of several years (Caves et al., 1991, Grabowski and Vernon, 1992, Hollis, 2002, Yu and Gupta, 2014, Shajarizadeh et al., 2015). For instance, as noted in Hollis (2002), in the Canadian market, the first generic advantage arises due to patients' unwillingness to switch between generic medications, the search and persuasion costs on the part of doctors, and the additional administrative costs of pharmacies when stocking several identical generic drugs with no real monetary incentives due to reference pricing. In the US, due to the presence of insurance and tiered co-payment system, patients may have a large incentive to switch from the branded to a generic drug (in the presence of a generic, the branded drug would be on a higher co-payment tier). However, the co-payment between first vs later generics would be the same (typically around $10-20) and so consumers would not have an added incentive to switch from the first generic to other generics. Thus, it can mean a strong first mover advantage for the first generic. In terms of the model that follows, it can also mean that if due to the presence of insurance, the profits of the branded firm do not erode much beyond the first generic entrant, then the threat by the branded firm to be in a triopoly via an AG is more likely to be credible.[3] As our point of entry, we take the first mover advantage as given (rather than model it), and instead model its impact via differences in maximum willingness to pay for a product, leading to asymmetries in demand curves for differentiated products. Thus, the 'prize' of being the first generic is not just a legislative market exclusivity period where the first generic entrant can operate as a duopolist, but also the relative order of entry – the rewards for which (due to the first mover advantage) are recouped by the entrant in the current period, as well as in the post-patent period when there may be several generic firms.

Another ingredient in our model is the ability of the branded firm to launch an authorized generic to deter entry. Liang (1996) discusses the incentives in launching a pseudo-generic prior to patent expiration to obtain first mover advantage in the generic segment for greater post-patent total

[3]Nonetheless, as shown in Berndt and Aitken (2011) and FTC (2011) generic prices can fall further as a function of number of generic manufacturers. While branded prices may not necessarily initially fall (may even increase for the brand loyal price inelastic patients after first generic entry as discussed in Frank and Salkever (1992)), the share of the branded market can erode further in the face of falling generic price index if the insurance companies exert greater pressure via higher co-payments to switch away from the branded segment. In the latter case, we would have the original situation where the branded firm faces a trade off in duopoly vs triopoly rents.

profits. Similarly, Kamien and Zang (1999), Kong and Seldon (2004) and, Rodrigues (2006) model launch of generic by a branded firm prior to patent expiration as a way to capture first mover advantage (via a Stackelberg leadership model) to deter entry. Instead, we differentiate between first and second generic via an explicit first mover advantage that the branded firm can capture by launching its authorized generic prior to an independent generic entry. The ability to launch its own authorized generic in turn determines the amount of payment offered to reach a P4D deal, or deter it from contesting entry in the first place.

Several studies have empirically documented the impact branded manufacturers have when they launch their own generic or an authorized generic via a third party on independent generic entry. Hollis (2003) argues that authorized generics deter independent generic entry in intermediate sized markets (and "probably" in other markets as well) while Reiffen and Ward (2007) show that authorized generic entry may deter independent generic entry in small and intermediate sized markets only and raise the long run prices by 1-2%. Farrell et al. (2011) also confirm the deterrent effect of AGs on generic producers in their empirical study of the effects of AGs on the revenues of successful challengers. Berndt et al. (2007) argue that the effect of authorized entry on independent generic entry – and ultimately on consumer welfare – is likely to be small but still positive. However, Appelt (2015) reports that early authorized entry has no impact on the likelihood of generic entry.

3. MODEL

3.1. **Setup.** We begin by describing a typical P4D deal from the US which serves as a motivating example for our stylized model. Shire Pharmaceuticals introduced an extended release version of its ADHD drug called Adderall XR in 2001. Under the Hatch-Waxman terms it had exclusivity until April 2005 (initial exclusivity was until October 2004, but then received pediatric extensions). The underlying patents for the extended release version, unless invalidated, were effective until 2018. In November 2002, Barr laboratories filed an abbreviated new drug application (ANDA) which was followed by a second filing by IMPAX in November 2003. Patent litigation ensued, but Shire settled with both parties before any court outcome. Shire settled with IMPAX (the second filer) to enter the market no later than December 2010, but with a non-exclusive license. It also settled with Barr

laboratories (the first filer), which acknowledged that Shire's patents were valid and agreed to delay entry until April 1, 2009. At that point, Barr would enter with a 180-day exclusive license from Shire and pay royalties as a proportion of its profits from the sales of generic Adderall XR over the exclusivity period (Barr Laboratories, Inc., 2006). Per the terms of the agreement, Barr may be allowed to enter early if another party were to successfully launch a generic version of the drug (for instance by invalidating Shire's patent). As per the terms of the deal, Teva Pharamceuticals (which had acquired Barr laboratories in the meantime) started marketing generic version of Adderall XR in the US on April 2, 2009, and six months later IMPAX also entered the market. For additional examples, see Hemphill (2007).

Based on the example above, and market entry rules summarized in Appendix B.1, we propose a sequential game Γ with $J+1$ players that illustrates the essential elements of interactions between a brand name firm B (player 0), which is protected by a patent, and $J \geq 1$ potential generic challengers $(G_1, \ldots, G_J)$ (in the appendix we also discuss how the model can be adapted to the European market entry rules).

(1) There are two periods, period 1 which is pre-patent expiration, and period 2, which is post-patent expiration period.
(2) In period one, the J potential entrants can sequentially contest entry.[4] The branded firm can offer a payment to a challenger to stay out of the market during period one (a P4D deal), and guarantee the order of entry in the post-patent period, as long as the patent is not invalidated by another challenger (order of entry is not guaranteed if the patent is invalidated).[5]
(3) If at any stage a challenger (say the jth) does not accept a P4D deal and wins the court case (patent is invalidated), that challenger enters immediately in period one. However, the remaining $J-j$ entrants can only enter in the next period. This assumption follows from the explicit exclusivity rules in the US, but in a later section we relax the assumption of exclusivity for first successful challenger to no exclusivity for anyone, or even to exclusivity restricted to just the first filer.[6]

[4]It is technically possible to have multiple first filers and in the case of the US, all of these first filers would be entitled to a joint exclusivity period, i.e., a shared exclusivity, the rewards from which would be small if there are many first filers. While we model sequential challenges, in a later section we analyze outcomes when there is no exclusivity available to anyone, which may have similar outcomes as when there are multiple first filers sharing the generic market.
[5]For instance, the branded firm can allow a generic to use its own production facilities to achieve all regulatory market approval requirements and enter just before other generic firms enter in post patent period.
[6]The 180-day exclusivity is explicit in the US but even after the 180-day exclusivity ends, remaining generics do not necessarily enter the market immediately. In fact, a report by FTC (pp. 98-99 2011) shows that in markets with AGs, entry by later

(4) Additionally, if the jth firm wins the court case, the brand can opt to launch an authorized generic (AG), either itself at an additional cost θ and earn two profits from the brand and its generic product, or via any of the previously paid-off firms, in which case it earns profits from the brand plus a licensing fee L.[7] If the brand launches an AG, period one consists of a triopoly. In what follows we also assume that if the brand launches an AG externally, it is only via the first generic challenger.[8]

(5) Payoffs from the second period are discounted by common factor $\delta \in [0, 1]$. Further, in this period we assume a competitive oligopoly ensues among the $J + 1$ firms, and there are no licensing agreements, as the patent has expired. However, the profits and/or market shares are not equal as the order of entry matters, i.e., one of the generic products has a first mover advantage over the other generics. For the base case we assume that the second through the last generic entrants all earn the same profit (which is less than that of the first generic entrant).[9]

Based on the rules above, the game is as follows. The patent can be challenged in any of the Γ_j subgames by a generic challenger j. In the first subgame Γ_1, a generic firm G_1 can choose to stay out of the market, in which case the monopoly continues and the game ends, or it can challenge entry. If it contests entry, the brand makes an offer of X_1 to G_1 to stay out of the market. If the offer is rejected, litigation ensues. If it is not rejected, the process is repeated with the second challenger. The game is depicted in Figures (A-1) and (A-2) in the appendix for the special case when there are only two potential challengers ($J = 2$). The game and payoffs differ slightly for the first versus the second challenger, and hence we show these two cases explicitly, but the generalization to $J > 2$ challengers is similar to the second challenger case, and we discuss that later.

generics happens slowly over time, taking on average 36 months after initial generic entry to evolve to peak value of seven generics in such markets.

[7] As noted in FTC (2011, p. 17-18), generics can be launched by the branded firm itself (in-house) as 'pseudogenerics', or via third parties, but require expertise in generic marketing. This is because while brand name drugs are typically marketed to physicians and consumers emphasizing product differentiation, and securing placement on formularies, generic drugs are marketed to wholesalers and pharmacies on the basis of price, consistency of supply, and ability to offer a large portfolio of drugs, which is a different expertise. Acquiring this expertise can be expensive, and we model it as a cost for the branded firm. Further, even if the cost to acquire in house capability is low, if the originator acquires such a capability then in order to reach a P4D deal, the branded firm would have to agree not to launch the in-house AG during the exclusivity period. Such an offer to not launch an in-house AG itself carries risk of being scrutinized by FTC and potential additional litigation costs in the law suit brought by the competition authority. Alternatively, the same report also notes that a potential cost of launching an AG is the cannibalization of sales of its branded product. For instance, the brand name drug revenue is significantly smaller in markets with an AG compared to markets with independent generics only (FTC, 2011, p. 59-61). Thus, an alternative is to model reduced revenue for the branded firm due to cannibalization, which would also lower its profit. As a modeling choice we decided to keep it simple via a one time cost rather than a per unit loss in revenue due to cannibalization.

[8] This is a simplification but follows the example from Shire-Barr deal mentioned above. An alternative is to randomize.

[9] In an alternative set up, we relax this assumption and allow the successful generic firm to earn more than other generics if it enters in period one (i.e. to model an incumbency advantage).

Continuing with the example of just two potential challengers, we denote equilibrium profits due to the sales of the branded or generic drugs in any period by $\Pi_j^M, \Pi_j^{D\#}$ and $\Pi_j^{T\#}$ where M, D, and T stand for profits in monopoly, duopoly and triopoly market structures respectively, and the subscripts $j \in \{0, 1, 2\}$ are for the brand and first and second generic entrants. The superscript '#' is set to 1 or 0 to indicate whether an authorized generic has been launched or not either by the branded firm itself as self-AG or via one of the paid off generic firms in a P4D deal. Further, unless the branded firm has launched an in-house AG, the discounted profits from the second period will be given by $\delta\Pi_j^{T0}$ rather than $\delta\Pi_j^{T1}$ as there are no licensing fees in the post-patent period per rule five.

We assume that monopoly profits are greater than industry profits in a duopoly, which are in turn greater than industry profits in a triopoly. Further, profits are negatively correlated with entry order, and thus in a triopoly, the branded firm has the highest profits followed by those of the first and then the second generic entrant. Note that the jth generic challenger is not necessarily the same as jth entrant since a generic firm can choose to stay out of a market, and hence we denote the profits of the jth *player* by V_j. For example, suppose generic 1 has been paid off and agrees to stay out of the market, and generic 2 enters the market and duopoly ensues between the brand and the second generic firm. Then, the equilibrium profits for the three players in the first period would be given by $(V_0^{D0}, V_1^{D0}, V_2^{D0}) = (\Pi_0^{D0}, 0, \Pi_1^{D0})$. Similarly, L_j is the adjustment to the final payoffs of the jth player due to any licensing agreements for an AG and we use the notation $\widetilde{V}_j^{T1} = V_j^{T1} + \delta V_j^{T\#} + L_j$ to indicate sum of equilibrium profits from the two periods plus any licensing fee (note that we use the superscript $T1$ on the sum of profits even if the second period is not necessarily T1, as long as the first period is T1). Also, since we assume that if an AG is launched it is only via the first challenger, we can simplify the notation to $L_1 = -L_0$ and $L_2 = 0$.

If at any of the two stages the generic rejects the offer, the involved parties incur litigation costs of c_0 and c_j (to be paid at the end of Γ_j). We assume c_0 is sufficiently low for B to always prefer litigation over unopposed entry and the ensuing competition. The outcome of the litigation is modeled by the fictitious player (N, Nature), who decides randomly with probabilities $1 - \pi$ and π, respectively

whether the brand B is successful with its lawsuit over patent infringement or not, and where π proxies the strength of patent ($\pi = 0$ being a very strong patent, and $\pi = 1$ being a very weak patent).[10]

As shown in Figures (A-1) and (A-2), the brand firm has the option of launching an AG at several of its decision nodes. For convenience, we will denote the subgames that start at these nodes as $\Gamma_{j,y}$, where j denotes the challenger and $y = \{B, G\}$ denotes the relevant path of the game: $y = B$ if either the brand wins the case or if the generic stays out, and $y = G$ if the generic wins. Note also that in the first stage when G_1 is the current challenger, the branded firm has the option to launch AG itself, whereas in the later stages, the option to launch an AG is only via the first paid-off generic firm (per the rules of our model). Hence, the first P4D deal contains - unlike the successive P4D deals - an (implicit) option to become an AG producer.[11] Further, if the branded firm launches a generic itself, the firm incurs a fixed cost θ (see subgame $\Gamma_{1,G}$ after B looses the patent litigation) or $\delta\theta$ if the generic is launched in the second period when it does not loose the litigation, or no generic firm challenges patent validity (see subgame $\Gamma_{1,B}$). Also if a self-AG is not launched and the generic does not challenge (as in $\Gamma_{1,B}$), the order of entry between the two generics for the second period is randomized, and hence the profits are depicted as expected generic profits. If both generic challengers have accepted P4D payments, the game ends at the Γ_3 node with payoffs given by $(\Pi_0^M - X_1 - X_2, X_1, X_2) + \delta(\Pi_0^{T0}, \Pi_1^{T0}, \Pi_2^{T0})$ which is similar to $\Gamma_{2,B}$ with an adjustment of X_2 payment to the second challenger but not drawn in the figure (and in the post-patent period profits are given by $\delta\Pi_j^{T0}$ instead of $\delta\Pi_j^{T1}$ since there are no licensing fees per the earlier rule five).

The final payoff to a player along a path of the game Γ consists of the corresponding (continuation) profit in the ensuing market structure adjusted by the P4D payments and/or litigation costs received and/or paid along the path. Except for some specific values of the parameters (where players are

[10] An alternative is to allow π to change with each challenger and model π_j as the subjective assessment of strength of the patent, which both parties agree on. In that case, settlement with a challenger may indicate that the patent is weak and hence π_j may increase with j. For an application with such an externality, see Palikot and Pietola (2018).

[11] Note also that in $\Gamma_{1,G}$ if the brand launches its own AG in the first period, the first and second generics' profits in post patent period are set equal to $\delta\Pi_2^{T1}/2$, i.e., they split the profits associated with a third product in a triopoly as there is no incumbency advantage for generic one, even though it enters in first period. The alternative extreme would be to assign $\delta\Pi_2^{T1}$ to generic one and zero economic profit to generic two with similar adjustments in $\Gamma_{j,G}$ for $j > 1$ cases. We consider the outcomes from such incumbency advantage in a later section.

indifferent between some of the alternative actions), the finite game Γ has a unique subgame perfect equilibrium (SPE) that can be readily computed by backward induction. In particular, we can compute the minimum offer that G_j, $j = 1, 2$, will accept in the SPE from the condition,

$$u_j(\Gamma_{j+1}) + X_j = \pi u_j(\Gamma_{j,G}) + (1-\pi)u_j(\Gamma_{j,B}) - c_j, \tag{1}$$

where $u_j(\Gamma.)$ is the expected payoff to player j in the unique SPE of the $\Gamma.$ subgame. The condition (1) makes the (risk neutral) player G_j indifferent between accepting X_j - and getting the left hand side (lhs) of (1) - and rejecting it - and expecting the right hand side (rhs) of (1). The brand B (player 0) will make the offer X_j in equilibrium, whenever its expected SPE payoff $u_0(\Gamma_{j+1})$ after paying X_j (receiving X_j if it is negative) exceeds its expected payoff from the litigation, i.e., when,

$$u_0(\Gamma_{j+1}) - X_j > \pi u_0(\Gamma_{j,G}) + (1-\pi)u_0(\Gamma_{j,B}) - c_0. \tag{2}$$

By combining (1) and (2), we obtain the condition for an agreement in Γ_j and the implied P4D payment as stated in the next lemma.[12]

Lemma 1. A P4D deal between the brand and challenger G_j will hold if the total payoffs to the two parties is greater than the sum of expectation from litigation. Specifically,

$$u_0(\Gamma_{j+1}) + u_j(\Gamma_{j+1}) > \pi\Big(u_0(\Gamma_{j,G}) + u_j(\Gamma_{j,G})\Big) + (1-\pi)\Big(u_0(\Gamma_{j,B}) + u_j(\Gamma_{j,B})\Big) - c_0 - c_j. \tag{3}$$

Further, under a take-it-or-leave-it offer, the amount X_j for a P4D is such that the generic is indifferent between triggering litigation and accepting the payment, the condition for which is given in (1). (*Proof* in Appendix (A.2)).

In this game of perfect information, G_j is able to compute the condition (3) and, in case it is satisfied, the P4D payment is as given in equation (1). Hence, it can rationally decide whether to challenge B or not. The following gives a condition under which G_j challenges B or stays out of the market.

Lemma 2. A generic G_j challenges the brand firm in a SPE if its P4D payment plus future expected payoffs are greater than the outside option of staying out of the market. This condition is given by

$$X_j + u_j(\Gamma_{j+1}) > u_j(\Gamma_{j,B}), \tag{4}$$

where X_j is defined in (1) (*Proof* in Appendix (A.2)).

[12]Note that our model allows X_j to be negative, in which case it is not a 'reverse' payment, or the usual P4D deal, but rather a 'forward' payment. This can happen, for instance, if the profits for a generic from being the first generic in the post-patent period are large enough so that it makes a payment to reserve this position.

3.2. **Endogenous Licensing Fee.** If the branded firm launches an authorized generic, it would charge a licensing fee. The authorized generic, however, is only launched if it increases the profit of the branded firm relative to an alternative outcome, and also increases the profit of the generic. In our game tree described above, and with just two challengers, this would be in the subgame $\Gamma_{2,G}$, where the second challenger G_2 rejects payment X_2 to stay out of the market, and the court decides in favor of the generic. In this case, in period one the brand's options are either to earn Π_0^{D0} (duopoly with no AG) or to earn Π_0^{T1} (triopoly with an AG launched via G1) plus a licensing fee.

We set licensing fee as a take-it or leave-it offer. Specifically, we compute the fee as a solution to an asymmetric Nash bargaining problem, but set the bargaining parameter equal to one for the branded firm ($\rho = 1$). Thus they reach a fee schedule by giving the entire surplus from the launch to the branded firm (i.e., the branded firm has full bargaining power). Then by launching an AG (T1 configuration), the profits due to the sales of the branded drug are $V_0^{T1} = \Pi_0^{T1} + \delta\Pi_0^{T0}$, where the second part is from sales in the post-patent period, and similarly, those due to sales of the authorized generic are $V_1^{T1} = \Pi_1^{T1} + \delta\Pi_1^{T0}$ (post patent profits are $\delta\Pi_j^{T0}$ instead of $\delta\Pi_j^{T1}$ as there is no licensing fee in the second period). On the other hand, by not launching the AG, the profits for the two products are $V_0^{D0} = \Pi_0^{D0} + \delta\Pi_0^{T0}$ and $V_1^{D0} = 0 + \delta\Pi_2^{T0}$, respectively (see Figure A-2). Note that we are explicitly accounting for the entry order of the challengers, where the first paid-off challenger either makes a profit $\delta\Pi_1^{T0}$ or $\delta\Pi_2^{T0}$ in the post-patent period, depending on whether it was launched in period one or not.

Thus, in this subgame with just two challengers, the net surplus from launching an AG is $(\Pi_0^{T1} + \Pi_1^{T1} - \Pi_0^{D0}) + \delta(\Pi_1^{T0} - \Pi_2^{T0})$, where the second term in the parentheses is due to the relative gain in profits of the first challenger in the post-patent period due to entering first or entering second. Consequently for arbitrary bargaining power ρ, two period profits inclusive of a licensing fee for the three firms are (if an AG is launched post losing a court case),

$$\begin{aligned} \widetilde{V}_0^{T1} &= \Pi_0^{D0} + \rho(\Pi_0^{T1} + \Pi_1^{T1} - \Pi_0^{D0}) \quad &+\delta(\Pi_0^{T0} + \rho(\Pi_1^{T0} - \Pi_2^{T0})) \\ \widetilde{V}_1^{T1} &= (1-\rho)(\Pi_0^{T1} + \Pi_1^{T1} - \Pi_0^{D0}) \quad &+\delta(\Pi_2^{T0} + (1-\rho)(\Pi_1^{T0} - \Pi_2^{T0})) \\ \text{and } V_2^{T1} &= \Pi_2^{T1} \quad &+\delta\Pi_2^{T0} \end{aligned} \tag{5}$$

where $\widetilde{V}_j^{T1}$ as defined earlier, is the sum of profits from the two periods adjusted by the licensing fee and under take-it or leave-it, $\rho = 1$.

3.3. **Extension to J > 2 and Related Cases.** The game naturally extends to more than two challengers where Γ_j is repeated for $j \in \{2, \ldots, J\}$ with the only difference being that the Π_j payoffs will be based on oligopoly profits $\Pi_j^{N\#}$ rather than for instance by triopoly profits given by $\Pi_j^{T\#}$. Appendix (A.3) provides an example of profits under a oligopoly with $J+1$ firms along with the structure of Γ_j and where we maintain the assumption that if the brand reaches an agreement with all the challengers, then in the post patent period the order of entry is given by the order of challengers (i.e., first paid off firm gets the first mover advantage).

Observe that solving Γ_j, i.e., finding out whether G_j challenges B and computing X_j, requires the solution to the game Γ_{j+1} first. Hence, SPE payoffs in Γ_j and all payments $X_J, \ldots, X_j$ are found by a recursive computation that uses equation (1) and Lemma (2) at each step $J, \ldots, j$. For example, if this computation yields that the generics $G_J, \ldots, G_j$ challenge B and agree on the P4D payments $X_J, \ldots, X_j$, then the brand's expected SPE payoff in Γ_j is,

$$u_0(\Gamma_j) = u_0(\Gamma_{J+1}) - \sum\nolimits_{s=j}^{J} X_s, \tag{6}$$

where $u_0(\Gamma_{J+1})$ is the payoff to the brand B after the game ends with J P4D agreements. If all these P4D payments are positive, condition (2) for a fixed j will be eventually violated when the number of generics J is sufficiently large. In this case, B and G_j will go to court. On the other hand, a condition for a universal agreement on P4D deals is specified in the next proposition.

Proposition 1. The brand will pay P4D payments to all challengers in a SPE if for any given challenger G_j, the brands expected payoff from agreements with the current and subsequent challengers (net of payments) is greater than the expected payoff from triggering the litigation against challenger G_j. Specifically, B will agree in the SPE on the P4D payments $X_1, \ldots, X_J$ if for all $j = 1, ..., J$,

$$u_0(\Gamma_{J+1}) - \sum\nolimits_{s=j}^{J} X_s > \pi u_0(\Gamma_{j,G}) + (1-\pi) u_0(\Gamma_{j,B}) - c_0, \tag{7}$$

where X_j is defined in equation (1). (For *Proof* see Appendix (A.2)).

4. RESULTS

4.1. **Parametrization and Equilibrium Profits.** To get intuition into the way the game unfolds, we need to specify equilibrium profits under alternative market structures (monopoly, duopoly, triopoly, N-opoly) and where the order of entry establishes a first mover advantage. In Appendix (B.2) we derive profit functions from the canonical differentiated products demand model by Singh and Vives (1984) but tweak it to account for first mover advantage. Specifically, in their model, utility for a representative consumer is given by $U(\mathbf{q}) = \boldsymbol{\alpha}\mathbf{q} - \frac{1}{2}\mathbf{q}'\boldsymbol{\Sigma}\mathbf{q}$, where the matrix $\boldsymbol{\Sigma}$ captures substitutability between products, and the vector $\boldsymbol{\alpha}$ specifies maximum willingness-to-pay (WTP) for each product. The derived demand curves from the utility maximization problem are linear and functions of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$.[13]

Further, to model first mover advantage, we allow $\alpha_j(\kappa)$ to be different for each product i and to depend on a parameter κ, which adds a degree of vertical differentiation between brand and the first and second generic entrants. Thus let $\kappa \in [0, 1]$ be such that in a triopoly, $\kappa = 0$ implies WTP for first and second generic is the same i.e. $\alpha_0^T(0) > \alpha_1^T(0) = \alpha_2^T(0)$, and $\kappa = 1$ means that the first generic entrant has the maximum advantage relative to the second generic entrant, where we set it to be the same as that for the branded drug, i.e., $\alpha_0^T(1) = \alpha_1^T(1) > \alpha_2^T(1)$. Similarly, in a duopoly $\alpha_0^D(0) > \alpha_1^D(0)$ and $\alpha_0^D(1) = \alpha_1^D(1)$ capture the WTP differences for between the branded and generic with and without first mover advantage.[14] Finally, we also hold constant the market size across alternative market structures, i.e. potential number of patients is the same across monopoly, duopoly or triopoly, so no new patients are discovered if a generic enters the market (though the actual realized demand may be different due to different prices).

With this setup we can compare equilibrium outcomes (prices, quantities, and profits) across market structures, which we use to solve the game. Without any P4D deals, our differentiated products

[13]Thus, for instance, in a triopoly $\boldsymbol{\alpha} = (\alpha_0^T, \alpha_1^T, \alpha_2^T)$ and based on utility maximization, derived demand involves intercepts (a_0^T, a_1^T, a_2^T) and slope coefficients. Similarly, in a duopoly $\boldsymbol{\alpha} = (\alpha_0^D, \alpha_1^D)$ and $\boldsymbol{\alpha} = (\alpha_0^M)$ in a monopoly. Also, $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix which we parameterize with just two terms, β on the leading diagonal and γ as the term on off-diagonals and in the case of a duopoly, $\boldsymbol{\Sigma}$ is a two by two matrix with similar terms, while in the case of a monopoly, it is a scalar equal to β. Further details are in Appendix (B.2).

[14]An additional parameter λ in our specification sets the relative market size between the generic and branded segments of the market, and is also determined by the WTP for generics relative to that of the branded product. When $\lambda = 1$, the total size of the generic market is fixed and set equal to the branded market $(a_1^{(T)} + a_2^{(T)} = \lambda a_0^{(T)}, \lambda = 1)$.

demand model shows that (i) in a duopoly, a branded drug earns more than the first generic so $\Pi_0^D(\kappa) \geq \Pi_1^D(\kappa)$ and that $\partial\Pi_1^D/\partial\kappa > 0$ and (ii) in a triopoly, profits are ordered as $\Pi_0^T(\kappa) \geq \Pi_1^T(\kappa) \geq \Pi_2^T(\kappa)$ and where $\partial\Pi_1^T/\partial\kappa > 0$, and $\partial\Pi_2^T/\partial\kappa < 0$. For selected values of the parameters, graphs of prices, quantities, and profits are shown in Figure A-3. The important aspect of these graphs is the general monotonic increase/decrease of profits in κ, as they do not depend on the specific values of the parameters (changing the values of the parameters only changes the relative magnitudes but not the shapes).[15]

4.2. **Credible Threat.** To better understand the amount of payments in P4D deals and when these will lead to later generic challengers staying out of the market, Figure 1 plots the equilibrium profits for the branded firm as a function of first mover advantage (κ) for different market structures. Specifically, the figure shows the profits of the branded firm under (i) a monopoly, (ii) a T0 competitive triopoly (with no AGs), (iii) a T1 tripoloy when the first generic is an AG and the brand earns a licensing fee L, and (iv) a D0 competitive duopoly.

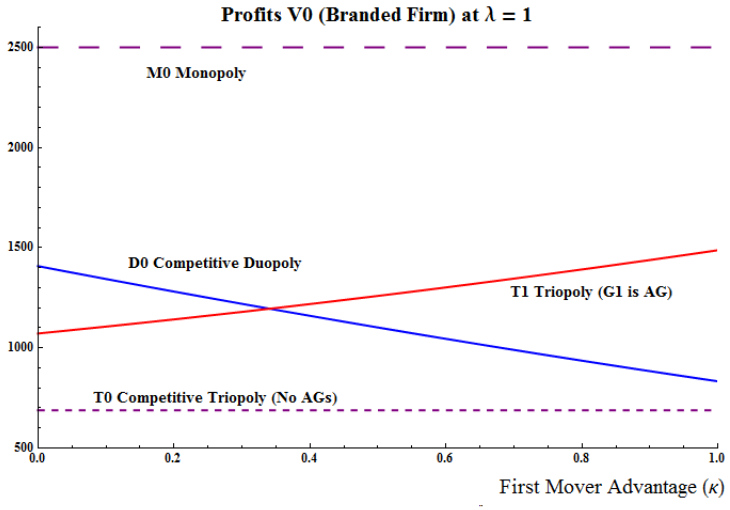


FIGURE 1. Profits of Branded Firm

The most desirable position from the perspective of the branded firm is the monopoly profit, and the least desirable is the competitive triopoly profit, neither of which change with κ (see Figure 1). Henceforth, for exposition, we will suppress κ when referring to functions $\widetilde{V}_j^{T\#}(\kappa), \widetilde{V}_j^{D\#}(\kappa), \Pi_j^{T\#}(\kappa)$ and $\Pi_j^{D\#}(\kappa)$. More generally when $\delta \neq 0$, note that $\widetilde{V}_0^{D0} = \Pi_0^{D0} + \delta\Pi_0^{T0}$ is decreasing in κ over the

[15] In these graphs, we set $\beta = 1$, $\gamma = 0.5$, $\alpha_0^M = 50/\gamma$, and $\lambda = 1$. Further, we set constant marginal costs to zero.

entire range while $\widetilde{V}_0^{T1}$, which is inclusive of the licensing fee, starts below $\widetilde{V}_0^{D0}$, but eventually is greater than $\widetilde{V}_0^{D0}$ (these are marked "*D0 Competitive Duopoly*" and "*T1 Triopoly (G1 is AG)*" in the figure). We label the intersection point of these two curves as the credible threat point κ^*, such that for all $\kappa \geq \kappa^*$, the branded firm's profits are higher in a triopoly with an authorized generic than in a competitive duopoly, i.e., $\widetilde{V}_0^{T1} \geq \widetilde{V}_0^{D0}$. The general shapes of these curves do not change much with the parameter values.[16] In the propositions below, we provide the condition under which a threat to launch an external AG becomes credible, and state the threshold value of θ above which the branded firm cannot launch its own generic.

Proposition 2. If in a T1 triopoly, the branded firm's profit are nearly constant in first mover advantage (i.e., $\partial\Pi_0^{T1}/\partial\kappa \approx 0$), and there exists a $\kappa^* \in [0,1]$ such that the net surplus from lunching AG at κ^* is zero, then under take-it or leave-it offer for the licensing fee, the threat to launch an AG is credible for all $\kappa \geq \kappa^*$. (*Proof* in Appendix (A.2)).

The condition $\partial\Pi_0^{T1}/\partial\kappa \approx 0$, that the equilibrium profit for the branded firm in T1 is nearly constant, is stronger than needed. What we need for net surplus to be increasing in κ is the condition $|\partial\Pi_0^{T1}/\partial\kappa| < |\partial\Pi_1^{T1}/\partial\kappa|$, i.e., the branded firm's equilibrium profit is decreasing in first mover advantage at a slower rate than the increase in the equilibrium profit of the first generic entrant so that the overall net surplus still keeps on increasing in κ (recall that Π_0^{T0} does not change with κ, but Π_0^{T1} can decrease in κ due to price coordination between the brand and the AG, see lower-left panel in Figure A-3 for the shape of Π_0^{T0}).

Next, we can provide conditions – or values of cost θ – under which a branded firm would prefer to launch an in-house AG. To do so, we fist define two threshold values. Let $\theta^*(\kappa) = (\Pi_0^{T1} + \Pi_1^{T1} - \Pi_0^{T0})$ and $\theta^{**}(\kappa) = (\Pi_0^{T1} + \Pi_1^{T0} - \Pi_0^{D0}) + \delta \cdot \theta^*(\kappa)$.

Proposition 3. If $\theta \leq \theta^{**}(\kappa)$ then branded firm would prefer to launch its own generic in period one if it loses litigation to the generic challenger. Alternatively, if it wins the challenge, it would prefer to launch in period two (post patent expiration) if $\theta \leq \theta^*(\kappa)$. (*Proof* in Appendix (A.2)).

[16] For instance, if the parameter λ mentioned earlier in footnote 14 is increased from 1 to 3, it increases the market share of generics relative to the brand and consequently moves the credible threat point to the left (graph omitted). Also, we have not shown the profit line for the branded firm when it can launch an in-house AG (so as not to clutter the graph). In fact for $\theta = 0$, it overlaps with the line "*T1 Triopoly (G1 is AG)*" since in the latter case the branded firm extracts all the surplus from the AG under the take-it or leave-it offer. For values of $\theta \geq 0$, the graph shifts to the right and for large enough values of θ the threat by a branded firm to launch its own generic may not be credible. Further, to launch own generic after loosing the case requires that the cost θ be lower than some threshold value ($\theta \leq \theta^{**}$), and that $\kappa \geq \kappa^*$.

Note that $\theta^{**}(\kappa)$ and $\theta^{*}(\kappa)$ are both increasing functions of κ, and if $\delta = 1$, then $\theta^{**}(\kappa)$ cuts $\theta^{*}(\kappa)$ from below (see Figure A-4). However if $\delta = 0$, then θ^{**} is always below θ^{*} (since $\Pi_0^{D0} \geq \Pi_0^{T0}$) and more importantly for low values of κ the threshold $\theta^{**}(\kappa)$ will be negative. This in turn implies that even if $\theta = 0$, for low values of κ, the brand will not find it profitable to launch its own generic and only credibly threaten to launch its generic against a winning challenger if the first mover advantage is large. We next discuss the payments to generic challengers.

4.3. **P4D Payments.** Consider first the subgame $\Gamma_{2,G}$ in which the first challenger (generic 1) has been paid-off an amount X_1 to drop the patent challenge, and in return will be allowed to enter first for the second period (providing the patent is not invalidated) while the second challenger (generic 2) is contesting the patent validity (see Figure A-2). If $\kappa \geq \kappa^*$, then the brand will always find it profitable to allow generic 1 to enter in period 1 as an AG rather than be in a competitive duopoly if the second challenger wins the court case.[17] Further, when the threat is credible, the second challenger's profits would be much lower than when it was not credible, i.e., they would be based on a triopoly with an AG (Π_2^{T1}) rather than on a competitive duopoly (Π_2^{D0}) as the branded firm can make sure that the AG enters first and claims the first mover advantage (see left panel of Figure A-5). In this case, generic 2 may well choose to stay out of the market and not challenge entry if its incremental expected profit post entry is less than its litigation costs, i.e., if $\pi\Pi_2^{T1} < c_2$. This is the incremental expected profit since in period two generic 2 would earn Π_2^{T0} either way.[18]

Alternatively, if the first mover advantage is not large, i.e. $\kappa < \kappa^*$, then the branded firm's preferred outcome is D0 duopoly over a T1 triopoly with an AG. In this case, the second challenger may well prefer to enter over the option of staying out since it can enter as a duopolist and grab the first mover advantage. However, a low value of κ also implies that the generic firm's profits are small as well. The brand has much to loose and will prefer to pay off the second challenger as well, and will pay $\pi\Pi_2^{D0} - c_2$, than lose its monopoly position.[19]

[17] Note that in this subgame, X_1 will be subtracted from both $\widetilde{V}_0^{D0}(\kappa)$ and $\widetilde{V}_0^{T1}(\kappa)$ hence the value of X_1 will not matter in the comparison.

[18] This situation is depicted in the right panel in Figure A-5 that shows the expected profits for the second generic for different values of π. The different values of π can be read as the strength of the patent, and when the litigation costs are set equal to 7.5% of the monopoly profits in this example.

[19] For the selected parameter values, at $\kappa = 0$, the generic could earn roughly 400 if it could invalidate the patent and enter, while the branded firms profit would drop from 2,500 in a monopoly to 1,400 in a duopoly.

Thus with just two potential challengers, either both generics will stay out of the market (an unchallenged monopoly) if the patent is strong (π is low) and/or cost of litigation is high, or the branded firm can always pay off both firms in P4D deals to maintain its monopoly in period 1. For a given litigation cost, whether it pays off both or only the first challenger, and the second optimally stays out, depends on π and κ with the possibility of paying off only the first firm starting at $\kappa \geq \kappa^*$. It can be verified that Proposition (1) implies the payment to the second challenger for a P4D deal (in the presence of J potential challengers) is

$$X_2 = \begin{cases} \pi\Pi_1^{D0} - c_2 + \delta\pi(\Pi_1^{T0} - \Pi_2^{T0}/(J-1)) & \text{if } \kappa < \kappa^* \\ \pi\Pi_2^{T1} - c_2 & \text{otherwise.} \end{cases} \tag{8}$$

Thus, for $\kappa \leq \kappa^*$ the challenger must be paid its expected profit in period one (as a duopolist) plus the expected premium due to the first mover advantage in period two minus the litigation costs, while if $\kappa \geq \kappa^*$, the payment to stay out falls to the expected profit of the second generic entrant in a triopoly minus the litigation costs. Note also the payment X_2 first increases in κ up to κ^*, and then falls in κ after that (see right side panel in Figure A-5 for graph of X_2 as a function of first mover advantage).

Similarly, the payments to the first challenger are also based on expected profits in period one (net of litigation costs) either as first entrant in a duopoly ($\pi\Pi_1^{D0} - c_1$) if $\theta > \theta^{**}(\kappa)$ or as a second entrant in a triopoly ($\pi\Pi_2^{T1} - c_1$) if $\theta \leq \theta^{**}(\kappa)$ plus the expected premium in period two, due to the first mover advantage. The exact values (given below) further depend on whether the branded firm can profitably launch an AG in the post patent period even if its generic does not enter in period one, i.e., if θ is less than or greater than $\theta^*(\kappa)$. Thus,

$$X_1 = \begin{cases} (\pi\Pi_1^{D0} - c_1) + \delta(1-\pi)[(\Pi_1^{T0} + \Pi_2^{T0})/J - \Pi_1^{T0}] & \text{if } \theta > \theta^{**} \text{ and } \theta > \theta^* \\ (\pi\Pi_1^{D0} - c_1) + \delta(1-\pi)[\Pi_2^{T1}/J - \Pi_1^{T0}] & \text{if } \theta > \theta^{**} \text{ and } \theta \leq \theta^* \\ (\pi\Pi_2^{T1} - c_1) + \delta(\pi\Pi_2^{T1}/J - \Pi_1^{T0}) + \delta(1-\pi)[(\Pi_1^{T0} + \Pi_2^{T0})/J] & \text{if } \theta \leq \theta^{**} \text{ and } \theta > \theta^* \\ (\pi\Pi_2^{T1} - c_1) + \delta(\Pi_2^{T1}/J - \Pi_1^{T0}) & \text{if } \theta \leq \theta^{**} \text{ and } \theta \leq \theta^* \end{cases} \tag{9}$$

where θ^* and θ^{**} are as defined earlier.

4.4. **Agreement Simulations.** We evaluated the game over combinations of κ and π values between zero and one and with alternative parameter values. Figure 2 shows the type of outcomes (litigation, P4D deals etc.) for four selected cases with parameter values given earlier but with litigation cost for all firms set equal to 7.5% of the monopoly profits. In the first panel, there are only two challengers and we have set both $\theta = 0$ and $\delta = 0$ ('θ = Low'). If the patent is strong ($\pi \approx 0$), the challengers choose to stay out (labelled 'I – Unchallenged Monopoly'). If the patent is weak ($\pi \approx 1$), the branded firm prefers to pay off the challengers and is able to do so rather than take its chances in a court (labelled 'II – P4D, Pay All'). Further, the boundary between both challengers being paid off choosing to not challenge the brand is marked by a tradeoff between the strength of the patent, and the relative first mover advantage: The actual payments to the two challengers are identical and increasing in κ until $\kappa = \kappa^*$, and are based on expected profits from entering in duopoly ($\pi\Pi_1^{D0} - c_j$). Starting at the threshold value, the payoffs to the challengers drop down to the expected profit of second generic entrant in a triopoly ($\pi\Pi_2^{T0} - c_j$) and thereafter further decrease in κ. Thus, as κ increases, the patent can be weaker and the two challengers can still be paid off by the generic firm.

An interesting case appears when the branded firm cannot launch its own generic because the cost to acquire generic marketing expertise is high. This is shown in the next panel (top right), marked as 'θ = High' (we set $\theta > \theta^{**}$ and $\theta > \theta^*$). As before, if the patent is strong ($\pi \approx 0$), neither generic challenges and the monopoly continues. For somewhat weaker patents, say $\pi \approx 0.6$, as we move in the direction of increasing the first mover advantage, level of payments and nature of P4D deals change. For $\kappa \leq \kappa^*$, the branded firm pays off both challengers and each is paid based on expected duopoly profits ($\pi\Pi_1^{D0} - c_j$). The magnitude of the payments becomes larger as κ increases. However, when $\kappa > \kappa^*$, the payments to the first challenger continue as before and keep increasing with κ, but the payments to the second challenger drop off to the level of second entrant in a triopoly ($\pi\Pi_2^{T0} - c_j$) and decrease with κ. This is because the second challenger can now be threatened with the launch of an AG via the first challenger. For a large enough value of κ (or equivalently for high litigation cost) the second challenger optimally stays out of the market (this area is labelled 'III – P4D, Pay Only First').

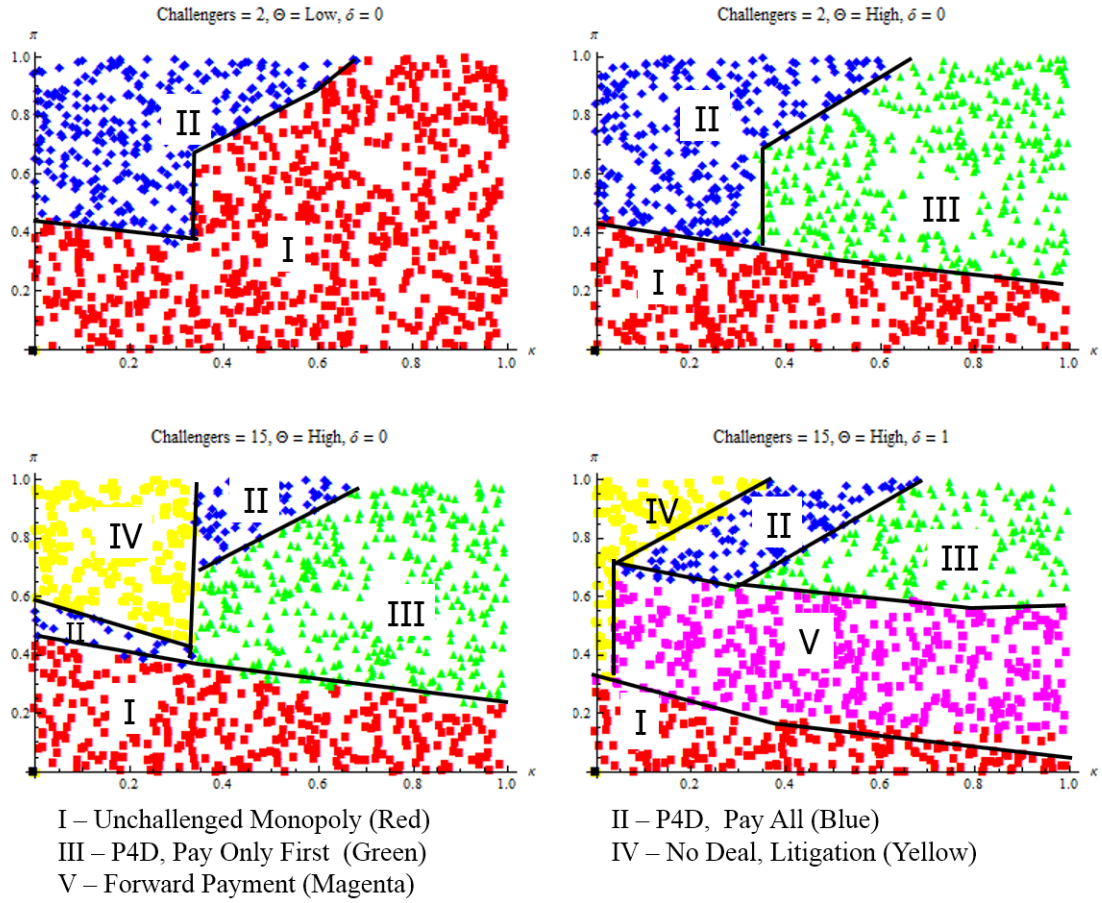


FIGURE 2. Agreement outcomes

When there are just two challengers then the branded firm can pay off both, but this is not always possible for a large number of challengers. This situation is depicted in the third panel (bottom left) for $J = 15$. When there are many potential challengers, the payments necessary to maintain the monopoly retain the form given above. Specifically, every challenger from the second one onwards must be paid off expected profits in D0 or T1 (depending on whether κ is less or equal to, or greater than κ^*) minus their litigation cost, and hence $X_j = X_2$ for $j = 3, \ldots, J$.

However, the net surplus with P4D deals with J challengers becomes negative making it impossible to pay off all the firms, i.e. condition in Proposition (2) is violated for large J. In this case, rather than paying off all the challengers, litigation ensues and the 'Pay All' region starts changing to 'IV – No Deal, Litigation' as shown in the third panel of Figure 2. The payments to later challengers are based either on what they would earn in a duopoly (i.e., are $\pi\Pi_1^{D0} - c_j$) if $\kappa < \kappa^*$, or are based on what they would earn as second entrants in a triopoly (i.e., $\pi\Pi_2^{T1} - c_j$) if $\kappa \geq \kappa^*$. Since duopoly

payments are larger than triopoly payments, the area to the left of κ^* converges faster in J to 'IV – No Deal, Litigation' compared to the area to the right of the credible threat point. However, as the number of challengers increases, all of the earlier 'Pay All' region becomes 'No Deal' region.

Observe also that increasing the number of challengers does not change the outcomes in the region earlier identified as 'III - P4D, Pay Only First'. Specifically, with a large number of challengers, the branded firm cannot afford to pay off all the firms. However, it can pay off the first challenger and the second onwards will not challenge as long as (i) $\kappa \geq \kappa^*$ and (ii) the patent is neither too strong (in which case no one challenges) nor too weak (where the brand anticipates a large number of small payments that exceed its ability to pay off and hence it does not offer P4D to any firm).

The last panel (bottom right) extends the forgoing analysis to the case when second period profits are also accounted ($\delta = 1$). While the payment formulas X_j are more complicated, the logic of agreement outcomes over the π, κ range is clear and the intuition is similar to when $\delta = 0$. Two main changes from the earlier case are that the threshold κ^* moves to the left, and that a new type of agreement outcome, 'V – Forward Payment' appears in the graph. The threshold moves to the left because $\widetilde{V}_0^{T1}$ has increased in magnitude more than $\widetilde{V}_0^{D0}$ (see equation (5)).

The new region is where the branded firm offers a *negative* payment to the first challenger to stay out of the market in period one and the challenger accepts this payment. We call this a forward payment region (opposite of 'reverse payment') because the generic firm makes a payment to the branded firm and stays out in the current period, but is able to enter first in the post patent period and grab the first mover advantage associated with its entry order. This payment goes to zero if either the future is discounted or if the branded firm has no ability to decide the order of entry. Since (in our model) the branded firm can always launch an AG just before the patent expiration to help a generic firm grab the first mover advantage, the firm is willing to pay to obtain that position.[20]

[20]In our model in case no generic firm challenges in period one, the order of entry among the J generics in the post patent period is randomly decided. In an alternative version of the model where we assign entry order in period two to be non-random and arbitrarily given to the first challenger, this new region never arises.

4.5. **Policy Option: No Exclusivity.** A popular policy option to discourage the pay-for-delay deals is to remove the six month exclusivity clause from the Hatch-Waxmax Act. Here we comment on the effectiveness of such a policy (equivalently, if we relaxed the exclusivity assumption in the model). If the jth challenger wins the court case, then sans the exclusivity period, all the remaining $J - j$ challengers can enter immediately in period one for free (i.e., without any litigation costs). Small changes in the payoffs in subgame $\Gamma_{j,G}$ accommodate this policy option and are given in Appendix (A.3).

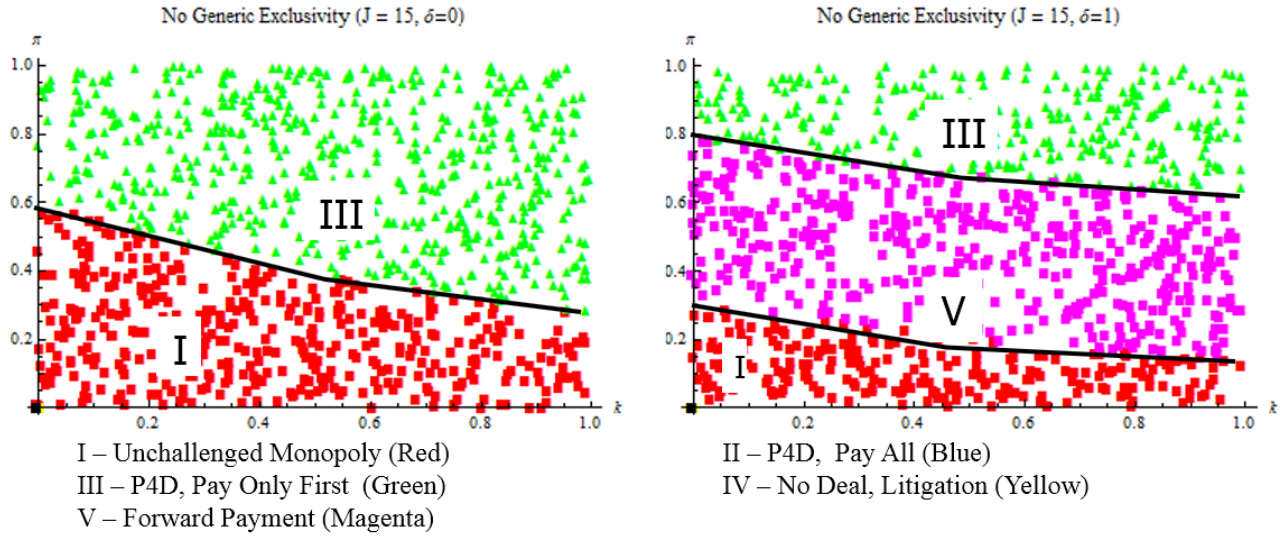


FIGURE 3. No Exclusivity

Since the expected profit of the challenger reduces from duopoly based rents to a competitive triopoly, this in turn lowers the payment required to keep the challenger out of the market. Similarly, if the branded firm does not launch an AG, its profits also decrease from Π_0^{D0} in period one to Π_0^{T0}. However, $\Pi_0^{T0} + \delta\Pi_0^{T0} \leq \widetilde{V}_0^{T1}$ for all values of κ even if it does not charge a licensing fee since it can coordinate on the price with an AG. Effectively, as before, the brand chooses between having one more firm that produces the drug as the first entrant AG with first mover advantage, or one less firm in an oligopoly but with no option to coordinate on price or charge a licensing fee. Consequently, the threat to launch an AG is credible for all values of κ and it is cheaper to pay off a challenger, making P4D deals still possible. The outcomes with $J = 15$ challengers and with $\delta = 0$ or $\delta = 1$ are shown in Figure 3. In both cases, P4D deals are still possible and in fact the area of unchallenged monopoly increases.

4.6. **Policy Option: No AG Against a Winning Challenger.** The branded firm's ability to credibly threaten to launch an AG in case a challenger wins a court case gives rise to the P4D deals. If this option is not available – and hence the threat is never credible – then with enough challengers in the market, a P4D deal will never be reached. In the US, this would mean amending the Hatch-Waxman Act so that it *also* applies to the branded firm: if no other generic firm can enter for 180 days as a reward for invalidating the patent, then the branded firm can also not launch an AG *prior to the exhaustion* of the 180-day exclusivity by a successful challenger. It is important to note that this policy is not saying that a branded firm cannot launch a generic (indeed prices fall once a generic enters, authorized or independent), only that the exclusivity under the Hatch-Waxman Act also applies to branded firm launching own generic, rather than only independent firms. To understand the implications of such a policy, with the same parameters as before, we modified the game imposing that the brand is (legislatively) prevented from launching an AG if a challenger is successful.

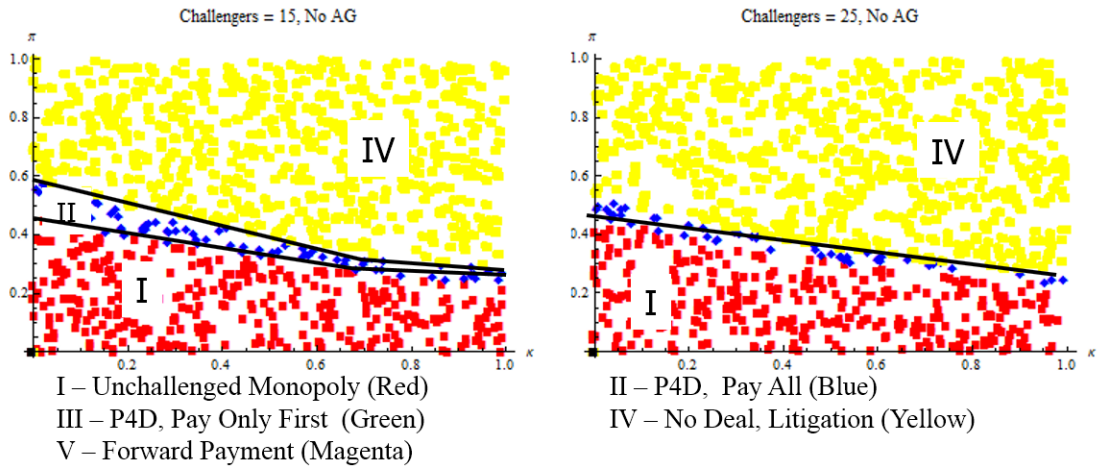


FIGURE 4. No Option to Launch AG

As shown in Figure 4, with no AG option against a successful challenger, either by itself or via a third party, the branded firm either has to pay off all the challengers (in case there are few challengers) or if there are many challengers, it may fail to reach an agreement with any of them. This is because after paying off the first challenger, the remaining $J-1$ challengers never optimally choose to stay out of the market, and hence the region marked as 'III – P4D Pay only First' never occurs. The only exception is when even the first firm does not consider challenging the branded

firm's patent because it is too strong ($\pi \approx 0$) relative to the litigation costs. All in all, removing the AG option for the brand leads to either an unchallenged monopoly for relatively strong patents, or a court decision rather than an out of court settlement if there are enough challengers.

4.7. **Robustness and Extensions.** Several other cases are relegated to the appendix but briefly discussed here. (1) We can incorporate the EU market entry rules in the model (see Appendix (A.3.3) for EU drug entry regulations). The main difference is that there is no explicit 180-days exclusivity period in Europe, and hence the outcomes would be similar to the ones depicted in the *No Exclusivity* case, i.e., an increase in parameter range over which pay-for-delay deals are possible. (2) We modified the model where exclusivity is made available to the first filer (FF) only instead of the first successful challenger (FSC). Once again the outcome is similar to the case of *No Exclusivity* as any later challenger cannot enjoy exclusivity benefits with the additional difference that the boundary between 'Unchallenged Monopoly' and 'Pay Only First Generic' shifts slightly downward (see Appendix (A.3.4)). (3) Our model is also robust to allowing for an incumbency advantage to the winning generic in post-patent period over other generics (see Appendix (A.3.5)). (4) Next, we also modified the payoff functions to allow for risk aversion by generic firms. Specifically, we modified the payoffs in the game tree to be exponential utility function of $\tilde{V}_j^{T\#}$ for generic firms. Details are given in the appendix, but the main differences in the agreement simulations are that while the threshold κ^* values does not change, the required payments to keep the challengers out of the market decrease, thus making it easier for the branded firm to pay off all challengers, and hence shrink the zone in which deals are rejected (area marked as 'IV No Deal, Litigation' shrinks). Additionally, ceteris paribus, generics were now also less likely to challenge monopoly position of a branded firm for the same underlying value of patent strength π, thus increasing the area marked as 'I - Unchallenged Monopoly' (see results in Appendix (A.3.6)). (5) Finally, we also compare the short term consumer surplus associated with various outcomes in the appendix (see Appendix (A.3.7)).

5. SUMMARY AND DISCUSSION

The model employed in this paper allows us to study the stability of reverse payment agreements between brand and generic challengers that lead to extended monopoly periods. Prior literature has focused on the welfare effects of out of court settlements with and without reverse payments, and under what conditions they may be anti or pro-competitive. We focused instead on when ex ante pay-for-delay deals would be observed in equilibrium, which is equivalent to exclusivity awarded to the first successful challenger, and how it compares to the current system in the US which awards exclusivity to the first filer.

Our model combines the first mover advantage for the first generic entrant with the ability of the branded manufacturer to launch an authorized generic to describe the conditions under which such deals are an equilibrium outcome. We do not explore all the other possible explanations for this phenomena. In particular, we show that compared to the first filer system, pay-for-delay deals are more difficult under the first successful challenger system. However even under the latter system, pay-for-delay deals can occur. We also show that pay-for-delay deals can occur even without any exclusivity period, as it happens in Europe.

The model also shows that the payment to stay out increases not only in the 'weakness' of the underlying patent, but also in the extent of the first mover advantage. This is important because both the US Supreme Court in the case against *Actavis*, and the European Commission (DG Competition) in announcing the €147m fine against Lundbeck and the agreeing generics in a pay-for-delay case, cite the size of the payment as a "workable surrogate" for the weakness of the underlying patent, but ignore the role of the first mover advantage.[21]

[21] See p.19 US Supreme Court (2013) and comments by the Director General (DG Competition) of EC, p.9 Italianer (2013).

REFERENCES

Appelt, Silvia, "Authorized generic entry prior to patent expiry: reassessing incentives for independent generic entry," *The Review of Economics and Statistics*, July 2015, *97* (3), 654–666.

Barr Laboratories, Inc., "Barr and Shire sign three agreements: a product acquisition agreement for Adderall; a product development agreement; and a settlement and license agreement for Adderall XR," PR Newswire [Press Release] August 14 2006. London, UK.

Berndt, Ernst R. and Murray L. Aitken, "Brand loyalty, generic entry and price competition in pharmaceuticals in the quarter century after the 1984 Waxman-Hatch legislation," *International Journal of the Economics of Business*, July 2011, *18* (2), 177–201.

_____, Richard Mortimer, Ashoke Bhattacharjya, Andrew Parece, and Edward Tuttle, "Authorized generic drugs, price competition, and consumers' welfare," *Health Affairs*, May/June 2007, *26* (3), 790–799.

Bernheim, B. Douglas, "Strategic deterrence of sequential entry into an industry," *RAND Journal of Economics*, Spring 1984, *15* (1), 1–11.

Bulow, Jeremy, "The gaming of pharmaceutical patents," in Adam B. Jaffe, Scott Stern, and Josh Lerner, eds., *Innovation Policy and the Economy*, Vol. 4, Cambridge, MA: MIT Press, 2004, pp. 145–187.

Caves, Richard E., Michael D. Whinston, and Mark A. Hurwitz, "Patent expiration, entry, and competition in the U.S. pharmaceutical industry," *Brookings Papers on Economic Activity*, 1991, *1991*, 1–66.

Crane, Daniel A., "Exit payments in settlement of patent infringement lawsuits: Antitrust rules and economic implications," *Florida Law Review*, 2002, *54* (4), 747–797.

Dickey, Bret, Jonathan Orszag, and Laura Tyson, "An economic assessment of patent settlements in the pharmaceutical industry," *Annals of Health Law*, Winter 2010, *19* (2), 367–400.

Drake, Keith M., Martha A. Starr, and Thomas G. McGuire, "Do "reverse payment" settlements constitute an anticompetitive pay-for-delay?," *International Journal of the Economics of Business*, 2015, *22* (2), 173–200.

EC, "Pharmaceutical Sector Inquiry, Final Report," European Commission (Directorates General, Competition) Brussels, Belgium July 2009.

Elhauge, Einer and Alex Krueger, "Solving the patent settlement puzzle," *Texas Law Review*, December 2012, *91* (283).

Eswaran, Mukesh, "Cross-licensing of competing patents as a facilitating device," *The Canadian Journal of Economics*, August 1994, *27* (3), 689–708.

Farrell, Joseph and Carl Shapiro, "How strong are weak patents?," *American Economic Review*, September 2008, *98* (4), 1347–1369.

_____, David J. Balan, Keith Brand, and Brett W. Wendling, "Economics at the FTC:

hospital mergers, authorized generic drugs, and consumer credit markets," *Review of Industrial Economics*, December 2011, *39* (4), 271–296.

Frank, Richard G., "The ongoing regulation of generic drugs," *The New England Journal of Medicine*, November 15 2007, *357* (20), 1993–1996.

_____ **and David S. Salkever**, "Pricing patent loss and the market for pharmaceuticals," *Southern Economic Journal*, October 1992, *59* (2), 165–179.

FTC, "Authorized generic drugs: Short-term effects and long-term impact," Federal Trade Commission Report Washington, D.C. August 2011.

Gallini, Nancy T., "Deterrence by market sharing: A strategic incentive for licensing," *American Economic Review*, December 1984, *74* (5), 931–941.

Glowicka, Elzbieta, Szabolcs Lorincz, Enrico Pesaresi, Lluis Sauri Romero, and Vincent Verouden, "Generic entry in prescription medicines in the EU: main characteristics, determinants and effects," July 8 2009. http://ec.europa.eu/dgs/competition/economist/prescription_medicines.pdf.

Grabowski, Henry G. and John M. Vernon, "Brand loyalty, entry and price competition in pharmaceuticals after the 1984 drug act," *Journal of Law and Economics*, October 1992, *35* (2), 331–350.

Graham, Stuart J. H., Bronwyn H. Hall, Dietmar Harhoff, and David C. Mowery, "Post-issue patent "quality control": A comparative study of US patent re-examinations and European patent oppositions," NBER Working Paper 8807, National Bureau of Economic Research February 2002.

Gratz, Linda, "Economic analysis of pay-for-delay settlements and their legal ruling," Munich Discussion Paper 2012-6, University of Munich, Department of Economics 2012.

Gürkaynak, Gönenç, Ayşe Güner, and Janelle Filson, "The global reach of FTC v. Actavis - Will Europe differ from the US approach to pay-for-delay agreements?," *IIC - International Review of Intellectual Property and Competition Law*, March 2014, *45* (2), 128–160.

Hancher, Leigh, "The EU pharmaceuticals market: parameters and pathways," in Elias Mossialos, Govin Permanand, Rita Baeten, and Tamara K. Hervey, eds., *Health systems governance in Europe: the role of European Union law and policy*, European Observatory on Health Systems and Policies, Cambridge: Cambridge University Press, 2010, chapter 15, pp. 635–682.

Harhoff, Dietmar, "Economic cost-benefit analysis of a unified and integrated European patent litigation system," Final Report MARKT/2008/06/D, Institute for Innovation Research, Technology Management and Entrepreneurship 2009.

Hemphill, C. Scott, "Drug patent settlements between rivals: A survey," Working Paper, New York University, School of Law March 12 2007.

_____, "An aggregate approach to antitrust: Using new data and rulemaking to preserve drug

competition," *Columbia Law Review*, May 2009, *109* (4), 629–688.

_____ **and Mark A. Lemley**, "Earning exclusivity: generic drug incentives and the Hatch-Waxman Act," *Antitrust Law Journal*, January 2011, *77* (3), 947–989.

Hollis, Aidan, "The importance of being first: evidence from Canadian generic pharmaceuticals," *Health Economics*, December 2002, *11* (8), 723–734.

_____, "The anti-competitive effects of brand-controlled "pseudo-generics" in the Canadian pharmaceutical market," *Canadian Public Policy - Analyse de Politiques*, 2003, *29* (1), 21–32.

Hovenkamp, Herbert J., Mark D. Janis, and Mark A. Lemley, "Anticompetitive settlement of Intellectual Property disputes," *Minnesota Law Review*, June 2003, *87* (6), 1719–1766.

Italianer, Alexander, "Competitor agreements under EU competition law," 40th Annual Conference on International Antitrust Law and Policy, Fordham Competition Law Institute, New York, 26 Septemeber 2013 2013. Comments by Director General for Competition, European Commission.

Kamien, Morton I and Israel Zang, "Virtual patent extension by cannibalization," *Southern Economic Journal*, July 1999, *66* (1), 117–131.

Kong, Ying and James R Seldon, "Pseudo-generic products and barriers to entry in pharmaceutical markets," *Review of Industrial Organization*, August 2004, *25* (1), 71–86.

Leffler, Keith and Cristofer Leffler, "Efficiency trade-offs in patent litigation settlements: Analysis gone astray," *University of San Francisco Law Review*, Fall 2004, *39*, 33–56.

Lemley, Mark A. and Carl Shapiro, "Probabilistic patents," *Journal of Economic Perspectives*, Spring 2005, *19* (2), 75–98.

Liang, Bryan A, "The anticompetitive nature of brand-name firm introduction of generics before patent expiration," *The Antitrust Bulletin*, 1996, *41*, 599.

Marxen, Annabelle and João Montez, "Generic entry in the pharmaceutical market: why less is better," mimeo 2018.

McGuire, Thomas, Keith Drake, Einer Elhauge, Raymond Hartman, and Martha Starr, "Resolving reverse-payment settlements with the smoking gun of stock price movements," *Iowa Law Review*, 2016, *81* (4), 1581–1599.

Mulcahy, Andrew, "Patent policy and entry: Evidence from pharmaceutical patent challenges." PhD dissertation, University of Pennsylvania, Philadelphia, PA January 2011.

Nevo, Aviv, "Identification of the oligopoly solution concept in a differentiated-products industry," *Economics Letters*, 1998, *59*, 391–395.

Padilla, Jorge and Valerie Meunier, "Should reverse payment patent settlements be prohibited per se?," Available at SSRN: https://ssrn.com/abstract=2604071 2015.

Palikot, Emil and Matias Pietola, "Pay-for-Delay with settlement externalities," Available at http://www.pietola.com/research/documents/patent2016.pdf 2018.

Regibeau, Pierre, ""Pay-for-Delay": What do we disagree on?," *Competition Policy International*, Autumn 2013, *9* (2), 116–127.

Reiffen, David and Michael R. Ward, "'Branded Generics' as a strategy to limit cannibalization of pharmaceutical markets," *Managerial and Decision Economics*, 2007, *28* (4-5), 251–265.

Rockett, Katharine E., "Choosing the competition and patent licensing," *RAND Journal of Economics*, Spring 1990, *21* (1), 161–171.

Rodrigues, Vasco, "Pseudo-generic products and barriers to entry in pharmaceutical markets: comment," *Review of Industrial Organization*, March 2006, *28* (2), 183–187.

Scott Morton, Fiona and Margaret Kyle, "Markets for pharmaceutical products," in Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds., *Handbook of Health Economics*, Vol. 2, Amsterdam, North-Holland: Elsevier Science Inc., 2011, chapter 12, pp. 763–823.

Scott Morton, Fiona M., "Pay-for-delay," *Competition Policy International*, Autum 2013, *9* (2), 128–136.

Shajarizadeh, Ali, Paul Grootendorst, and Aidan Hollis, "Newton's first law as applied to pharmacies: why entry order matters for generics," *International Journal of the Economics of Business*, June 2015, *22* (2), 201–217.

Shapiro, Carl, "Antitrust analysis of patent settlements between rivals," *Antitrust*, Summer 2003, *17*, 70–77.

———, "Antitrust limits to patent settlements," *RAND Journal of Economics*, Summer 2003, *34* (2), 391–411.

Singh, Nirvikar and Xavier Vives, "Price and quantity competition in a differentiated duopoly," *Rand Journal of Economics*, Winter 1984, *15* (4), 546–554.

US Supreme Court, "*FTC v. Actavis, Inc.,* 570 U.S. (2013), Slip Opinion," Supreme Court of the United States, No 12-416 2013.

Vives, Xavier, "Duopoly information equilibrium: Cournot and Bertrand," *Journal of Economic Theory*, October 1984, *34* (1), 71–94.

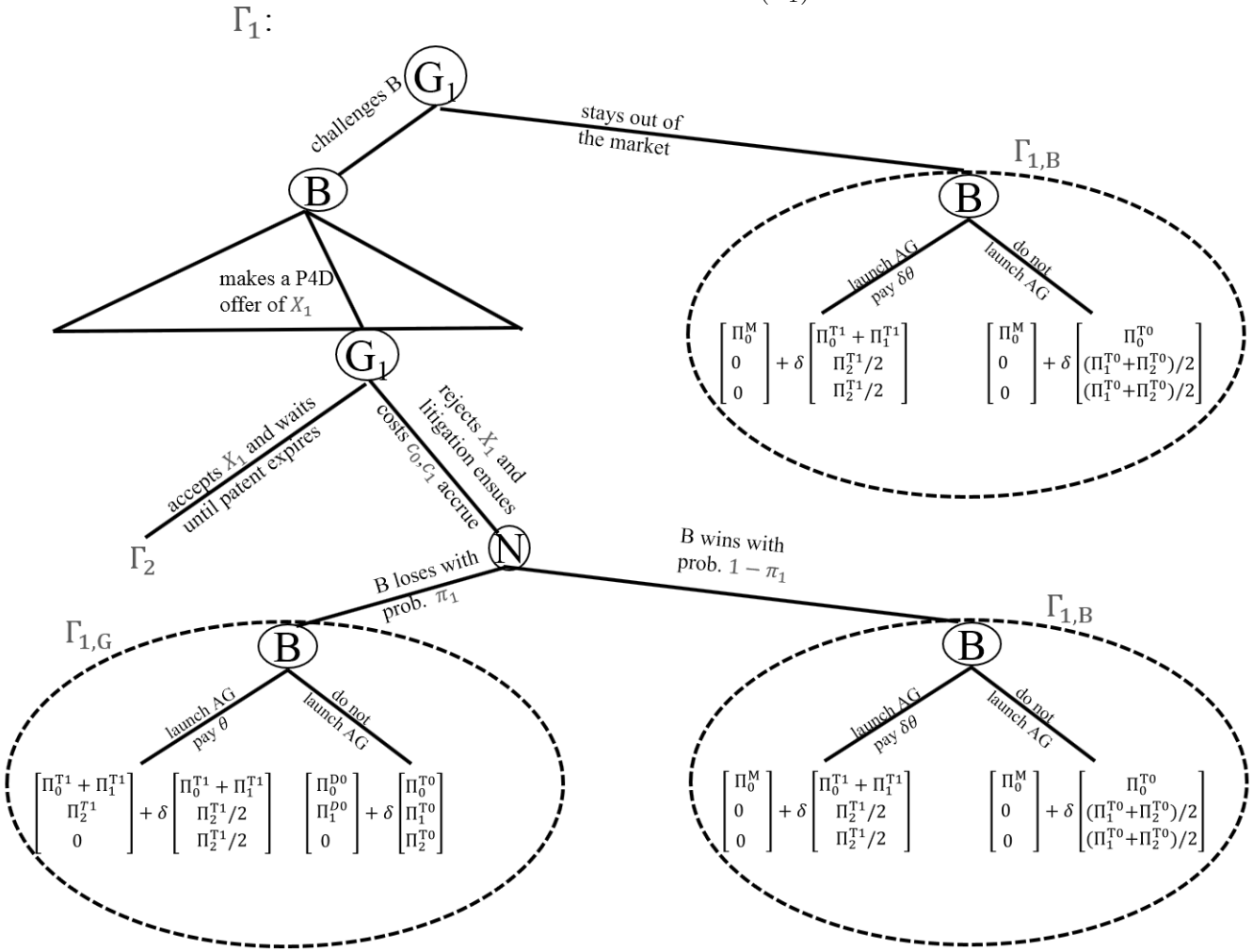
Willig, Robert D. and John P. Bigelow, "Antitrust policy toward agreements that settle patent litigation," *The Antritrust Bulletin*, Fall 2004, *49*, 655–698.

Yu, Yu and Sachin Gupta, "Pioneering advantage in generic drug competition," *International Journal of Pharmaceutical and Healthcare Marketing*, 2014, *8* (2), 126–150.

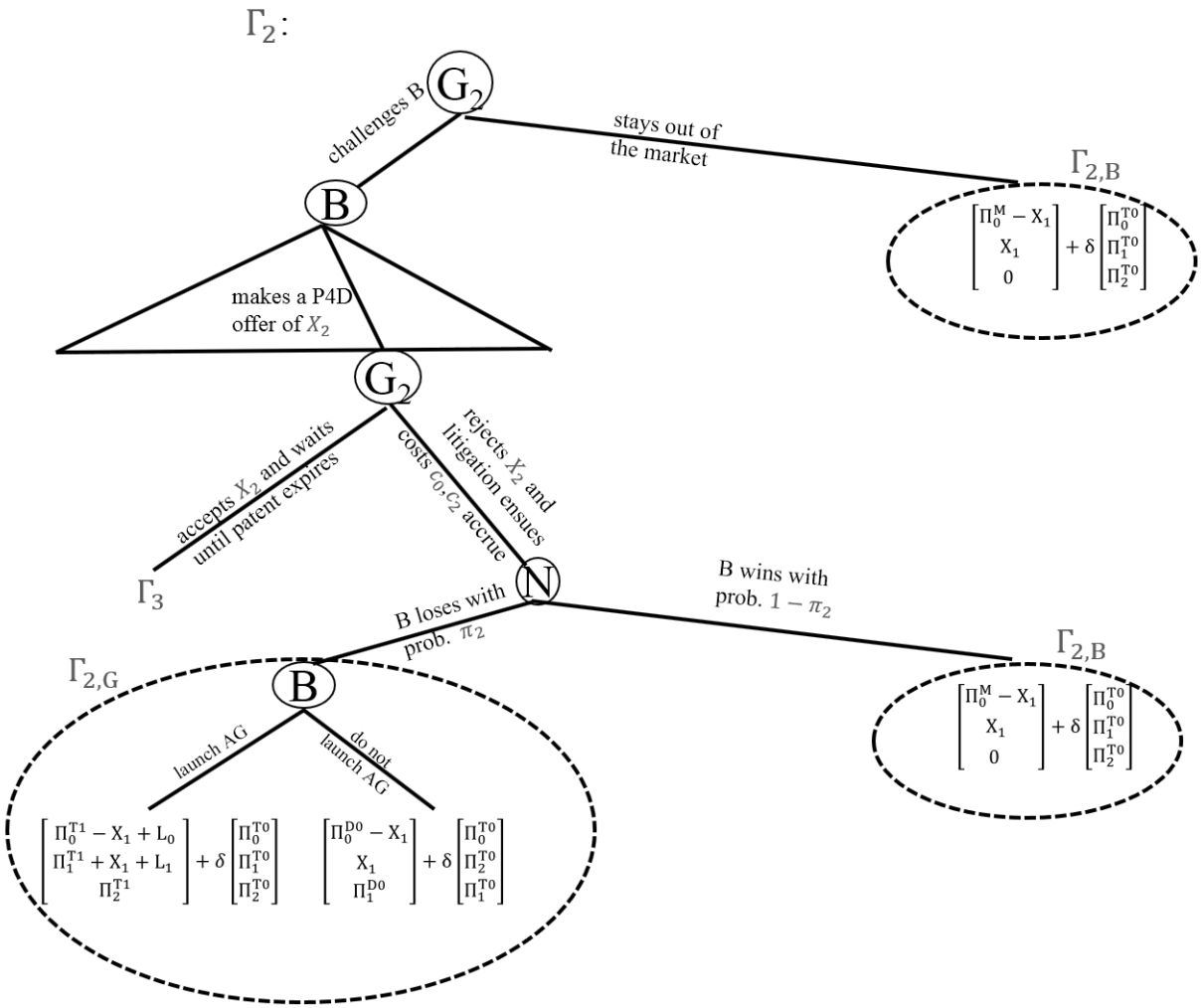
APPENDIX A. (FIGURES, PROOFS AND EXTENSIONS)

A.1. **Figures.** This appendix provides figures referred to in the paper.

FIGURE A-1. Game Tree (Γ_1)



Note that only equilibrium profits from sales are shown in the nodes. The final payoffs include also litigation costs and AG costs as indicated along the branches.



Note that only equilibrium profits from sales, P4D payments and licensing fees are shown in the nodes. The final payoffs also include litigation costs and AG costs as indicated along the branches.

FIGURE A-2. Game Tree (Γ_2)

FIGURE A-3. Non-Collusive Triopoly, Duopoly, and Monopoly

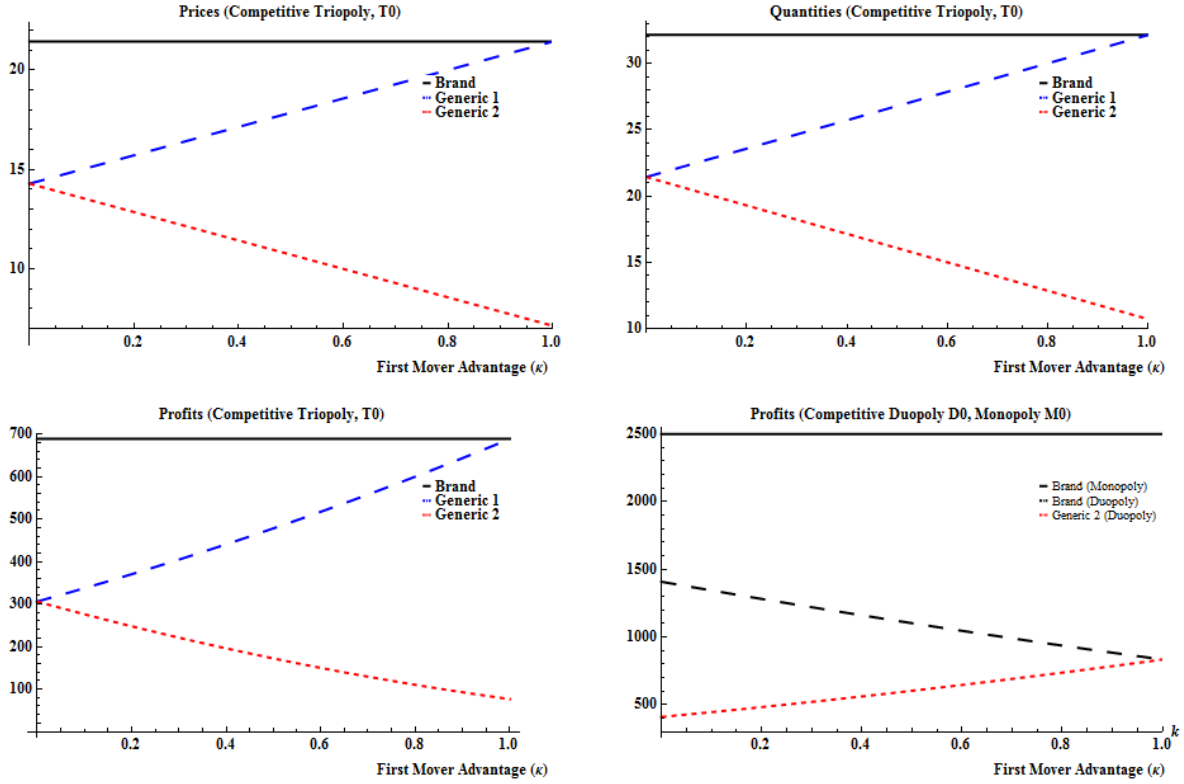


FIGURE A-4. Cost Thresholds

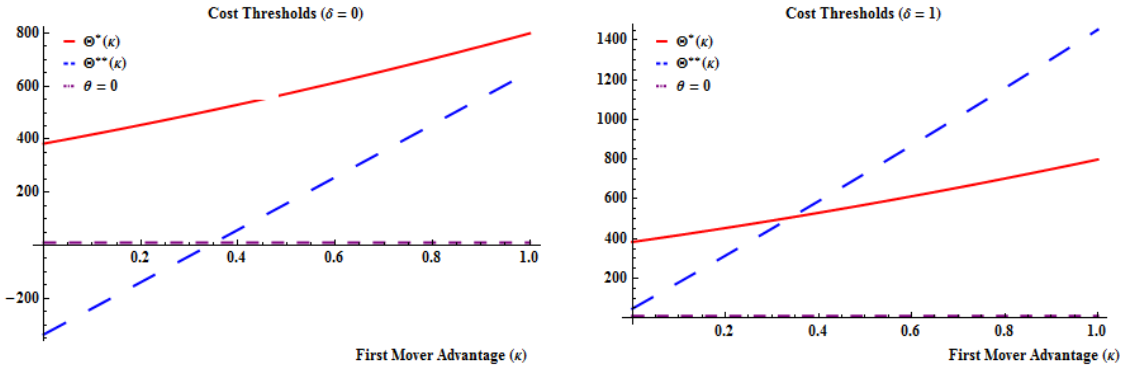
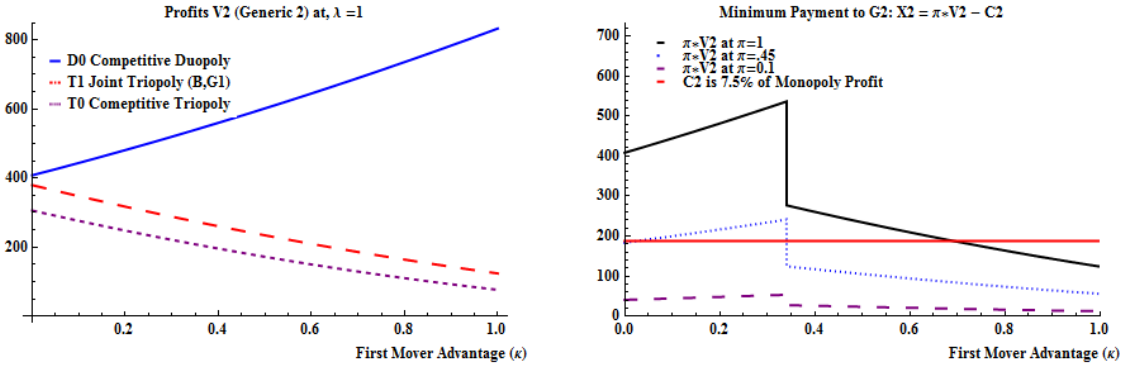


FIGURE A-5. Profits and Payment to Second Challenger



A.2. **Proofs.**

Proof of Lemma 1. The condition in the proposition obtains as the sum of (1) and (2). If this condition holds, then the net agreement surplus, i.e., the total continuation payoff to B and G_j after agreement minus their total payoff after disagreement, given by,

$$\Big(u_0(\Gamma_{j+1}) + u_j(\Gamma_{j+1})\Big) - \Big(\pi(u_0(\Gamma_{j,G}) + u_j(\Gamma_{j,G})) + (1-\pi)(u_0(\Gamma_{j,B}) + u_j(\Gamma_{j,B})) - c_0 - c_j\Big),$$

is positive, and both parties will rationally agree. As B makes a take-it-or-leave-it offer in Γ_j, it will extract the entire net surplus. This post-agreement sharing rule is implemented by the P4D payment (1). If the net surplus is negative, i.e., the condition in the proposition does not hold, B prefers the litigation to the agreement. Hence, an unacceptable offer (below X_j) will be made by B, rejected by G_j, and litigation will ensue. □

Proof of Lemma 2. After challenging B, the generic G_j expects the payoff $X_j + u_j(\Gamma_{j+1})$ in case of agreement with B. This amount is equal to the expected G_j's payoff after disagreement as the substitution from (1) shows,

$$X_j + u_j(\Gamma_{j+1}) = \pi u_j(\Gamma_{j,G}) + (1-\pi)u_j(\Gamma_{j,B}) - c_j.$$

Hence, G_j's expected payoff $X_j + u_j(\Gamma_{j+1})$ after challenging B does not depend on the outcome of the bargaining stage in Γ_j. On the other hand, if G_j does not challenge B, its continuation payoff is $u_j(\Gamma_{j,B})$. A rational G_j will challenge B if the former payoff is greater than the latter. □

Proof of Proposition 1. For each $j = 1, ..., J$, the left hand side of (7) is the SPE payoff to B upon agreement with G_j in Γ_j and subsequent agreements with $G_{j+1}, ..., G_J$ (for $j < J$). Hence, B anticipates in Γ_j that it will make equilibrium P4D payments to G_j and all subsequent challengers if (7) holds for $j, ..., J$. The rhs of (7) is B's expected payoff from litigating G_j (and avoiding the payments $X_j, \ldots, X_J$). Hence, B will agree with all challengers if the former payoff is greater than the latter for all $j = 1, ..., J$. □

Proof of Proposition 2. The net surplus from launching an AG with a previously paid off firm when the brand has lost the litigation to a challenger is $(\Pi_0^{T1} + \Pi_1^{T1} - \Pi_0^{D0}) + \delta(\Pi_1^{T0} - \Pi_2^{T0})$. Note that as long as $\partial\Pi_0^{T1}/\partial\kappa \approx 0$, the net surplus is increasing in κ: $(\Pi_1^{T0} \geq \Pi_2^{T0})$ for all values of κ with equality only when there is no first mover advantage, i.e., $\kappa = 0$, while Π_1^{T1} and Π_0^{D0} are respectively monotonically increasing and decreasing in κ. Thus, with the net surplus equal to zero at κ^*, it is positive for all $\kappa > \kappa^*$ and hence the threat is credible for $\kappa \geq \kappa^*$. □

Proof of Proposition 3. The first part follows directly from the subgame $\Gamma_{1,G}$ (see payoffs given in Figure A-1). After loosing to G_1, brand would launch a generic if $\Pi_0^{T1} + \Pi_1^{T1} - \theta + \delta(\Pi_0^{T1} + \Pi_1^{T1}) \geq \Pi_0^{D0} + \delta\Pi_0^{T0}$. Rearranging the terms gives the required result $\theta \leq (\Pi_0^{T1} + \Pi_1^{T0} - \Pi_0^{D0}) + \delta \cdot (\Pi_0^{T1} + \Pi_1^{T1} - \Pi_0^{T0}) = \theta^{**}(\kappa)$. Similarly, the second part follows from the subgame $\Gamma_{1,B}$. After winning against G_1, the brand launches a generic in post-patent period if $\Pi_0^M + \delta(\Pi_0^{T1} + \Pi_1^{T1} - \theta) \geq \Pi_0^M + \delta(\Pi_0^{T0})$. Rearranging gives the required result $\theta \leq (\Pi_0^{T1} + \Pi_1^{T1} - \Pi_0^{T0}) = \theta^*(\kappa)$. □

A.3. **Extension to the Game Tree.**

A.3.1. *Payoffs with $J > 2$ Firms.* In the game with $J > 2$ challengers, let the equilibrium profits of the jth player from sales of its product be given by $\Pi_j^{N\#}$ (see Figure A-6). We model these similarly to those in the triopoly where the first two players earn profits equal to that of the brand and the first generic in a triopoly, and all the later entrants equally share profits associated with the third player in a triopoly (an alternative is to set the profits of later entrants to zero which did not change our results in any significant way). Thus, for instance, in the post patent period with no AGs, the profits would be given by $(\Pi_0^{T0}, \Pi_1^{T0}, \Pi_2^{T0}/(J-1), \ldots, \Pi_2^{T0}/(J-1))$ and hence the final payoffs are accounted using the values $\Pi_j^{T\#}$ depending on the entry order. Then the Γ_j subgame would be as shown in the figure below.

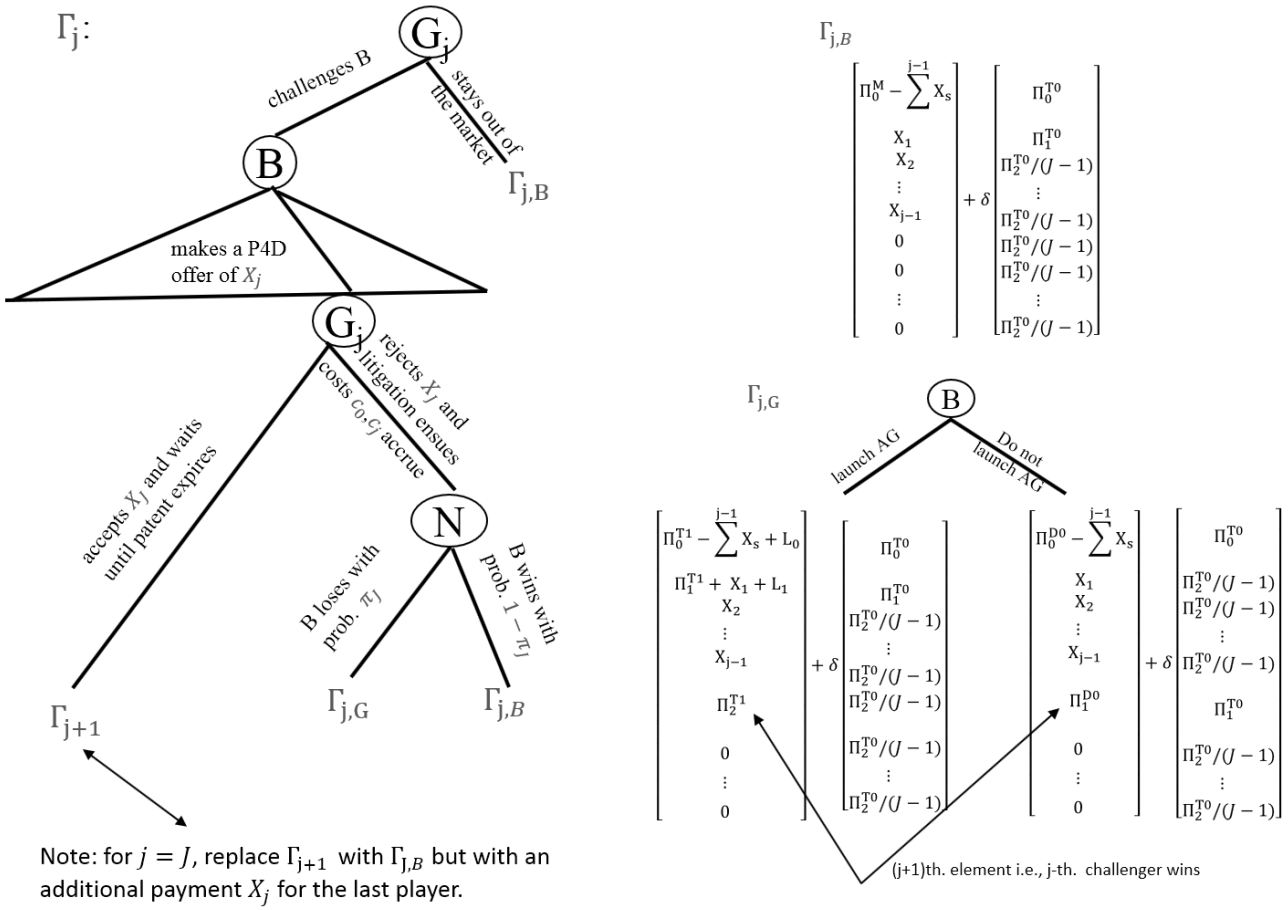


FIGURE A-6. Game Tree (Γ_j) with $J > 2$ players

Note that if B looses to the j-th challenger ($j > 1$), then the choice to launch AG or not in the $\Gamma_{j,G}$ is the same as before. Further, if AG is launched, the first mover advantage does not go to the winning challenger. The latter earns Π_2^{T1} in the current period and $\delta\Pi_2^{T0}/(J-1)$ in period 2, while if the AG is not launched, it earns a duopoly profit in the current period and grabs the first mover advantage earning Π_1^{D0} and $\delta\Pi_1^{T0}$ in the first and second periods respectively.

A.3.2. *No Exclusivity.* Consider the payoffs if the jth challenger wins the court case and all the remaining $J-j$ challengers can enter immediately in period one for free (i.e., without any litigation costs). Then building on our earlier specification where the profits for firms can be approximated as

in a triopoly (the brand and the first entrant earn profits of the first two firms in a triopoly $\Pi_0^{T\#}$ or $\Pi_1^{T\#}$ and the profit of all the remaining entrants is equal to the profit of the third firm in a triopoly divided by the number of $J-j$ remaining entrants $\Pi_2^{T\#}/(J-j)$), only the payoffs in the subgame Γ_{jG} change.

The payoffs in the $\Gamma_{j,G}$ subgame are modified as shown in Figure A-7 below for the case when a win by the jth generic implies that all the remaining $J-j$ potential challengers enter in the current period. Specifically, if the brand does not launch an AG but all other challengers can enter in period one, the potential profits for the winning jth challenger change from $(\Pi_1^{D0}+\delta\Pi_1^{T0})$ to $(\Pi_1^{T0}+\delta\Pi_1^{T0})$, while if an AG is launched, they change from $(\Pi_2^{T1}+\delta\Pi_2^{T0}/(J-1))$ to $(\Pi_2^{T1}/(J-j+1)+\Pi_2^{T0}/(J-1))$. The remaining challengers also earn positive amounts rather than zero in the first period (see tree below).

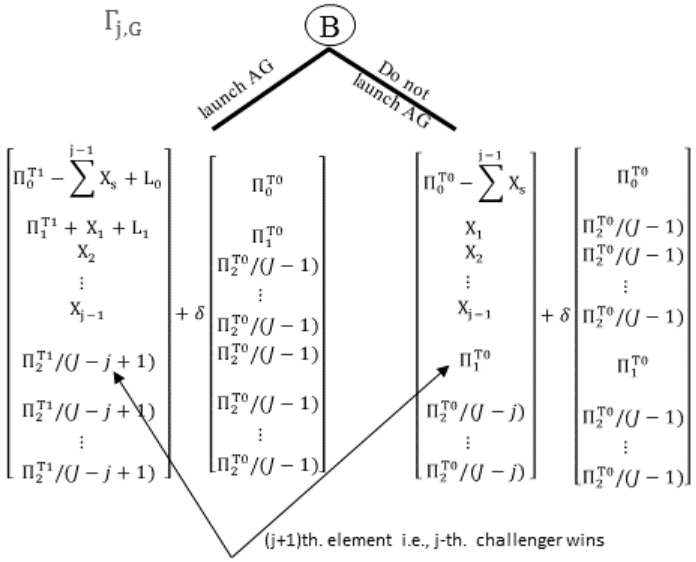


FIGURE A-7. $\Gamma_{j,G}$ Under No Exclusivity

A.3.3. *Extension to the European Market.* The 180-day exclusivity is explicit in the US but not in the EU. Nonetheless, in the EU there are other barriers to entry, such as delays in market authorization by the medical agencies that create de facto duopoly periods for the first successful generic challenger, making it perhaps similar to the American case. These delays which are sometimes caused by too many applications being filed with the EMA may be strategic and were noted in the *Pharmaceutical Sector Inquiry* by DG Comp (EC, 2009). The outcomes in that respect should be similar to the ones shown in the main document for the American case. An alternative however is to model the European case as when all the remaining generics enter the market as soon as one generic invalidates the patent. In that case, the European case would mimic the *No Exclusivity* option modeled above and for which results are shown in the main text.

A.3.4. *First Filer (FF) Exclusivity.* In the main model, we have assumed that the exclusivity is made available to the first successful challenger (FSC) rather than the first filer (FF). We now consider the case when exclusivity is available to only first filer to validate our claim that under such a system, pay-for-delay deals are permissible under a much larger range of model parameters.

In this case if the jth firm wins the litigation, its entry is still blocked if AG is launched (since the first generic has the exclusivity). Thus if the AG is launched the profit for the first generic changes

from $(\Pi_1^{T1} + \delta\Pi_1^{T0})$ to $(\Pi_1^{D1} + \delta\Pi_1^{T0})$ plus X_1 payment and minus licensing fee in both cases while the payoff for the winning challenger changes from $(\Pi_2^{T1} + \delta\Pi_2^{T0}/(J-1))$ to $(0 + \delta\Pi_2^{T0}/(J-1))$. If however the AG is not launched (post a win by the jth generic) then exclusivity is lost to all due to the forfeiture clause in Medicare Modernization Act of 2003, and the winning challenger and remaining firms enter immediately. In this case the payoff for the jth firm is $\Pi_1^{T0} + \delta\Pi_1^{T0}$ and of the remaining generics without P4D deals is $\Pi_2^{T0}/(J-j+1) + \delta\Pi_2^{T0}/(J-1)$.

The payoffs in the $\Gamma_{j,G}$ subgame are modified as shown in Figure A-8 below for the case when a win by the jth generic implies that either the exclusivity is available to only the first generic (where it is launched as an AG) and the winning jth generic cannot enter in period one, or exclusivity is not available to anyone if the first generic does not enter (forfeiture clause), in which case the jth challenger and all the remaining firms can enter immediately.

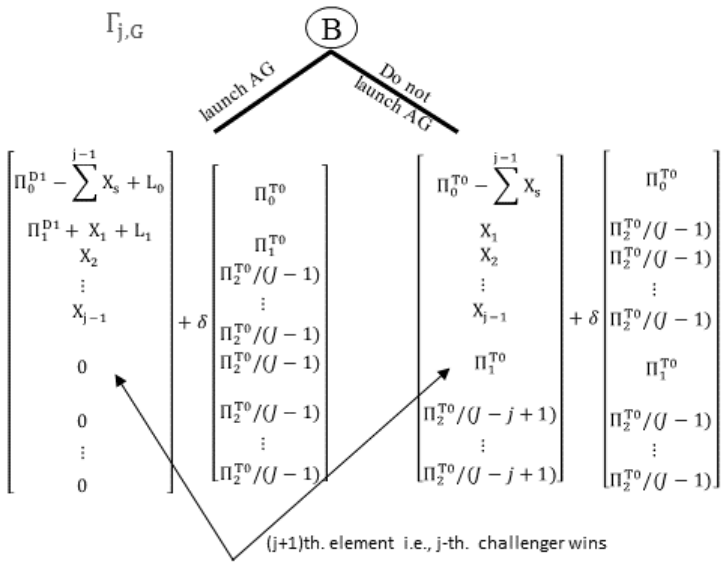


FIGURE A-8. $\Gamma_{j,G}$ Under FF Exclusivity

The outcomes (and logic) in this case is similar to the *No Exclusivity* case, where the expected profit of the jth challenger reduces from $(\Pi_2^{T1} + \delta\Pi_2^{T0}/(J-1))$ to $(0 + \delta\Pi_2^{T0}/(J-1))$ if an AG is launched making it easier to pay off this challenger. The only difference we observe in the corresponding figures is that boundaries between the regions shown in Figure 3 shift downwards, indicating that P4D deals are available over a larger parameter space (and hence for brevity the figure is omitted).

A.3.5. *Incumbency Advantage.* In the FSC system discussed in the paper, the second challenger, were it to enter successfully in period one, does not have an advantage over other generics in post patent period who do not enter in the first period. For instance, in our payoff specification above, if the j-th generic wins the litigation and the brand launches AG (say via the first challenger), then the winning generic earns $\delta(\Pi_2^{T0}/(J-1))$ in the post patent period, which is the same as what other non-entering generics earn in the second period (and the AG earns $\delta\Pi_1^{T0}$). An alternative is that the winning generic earns more than other generics who do not enter, and in an extreme case captures the entire generic residual market. This may be an important factor for some drugs.[22] Thus, we consider the other extreme where the incumbency advantage is at its maximum and the

[22]A report by FTC (2011, see graph on p.104) shows pronounced market shares for second generic (in their graph first-filer) relative to late entrants in post-exclusivity period when the AG and first-filer both enter in the first period.

winning jth generic earns all of the $\delta\Pi_2^{T0}$ in the post-patent period if it enters in period one, while other generics earn zero profits. This requires that the payoffs in $\Gamma_{j,G}$ are adjusted accordingly as shown in Figure A-9 below for the case when a win by the jth generic implies that if it enters in the current period (after winning the case), it will have an advantage over other generics in the post-patent period.

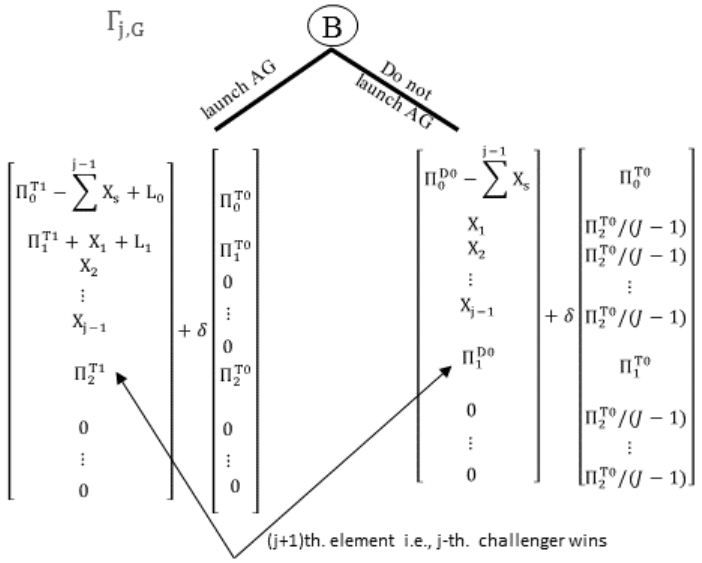
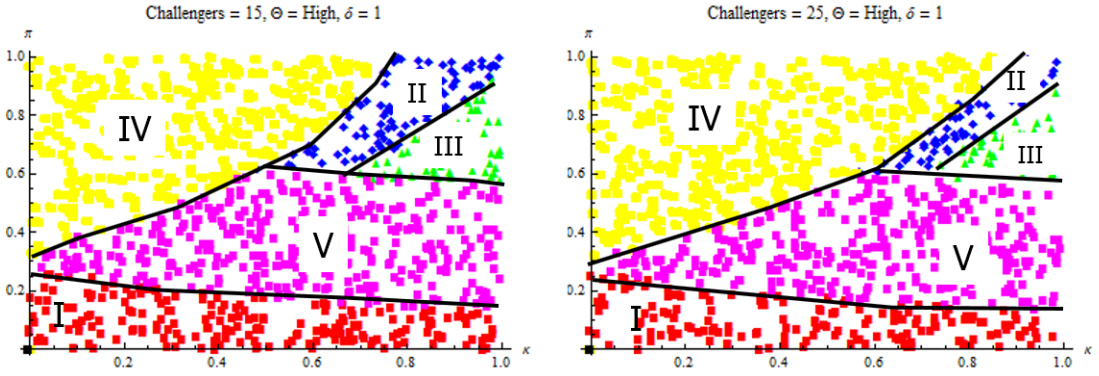


FIGURE A-9. $\Gamma_{j,G}$ with an Incumbency Advantage



I – Unchallenged Monopoly (Red)
II – P4D, Pay All (Blue)
III – P4D, Pay Only First (Green)
IV – No Deal, Litigation (Yellow)
V – Forward Payment (Magenta)

FIGURE A-10. Agreement outcomes with Incumbency Advantage

With this change in payoffs, note that the outcomes depicted for the first three panels in Figure 2 do not change (because $\delta = 0$ in these cases). However, the payments X_j increase slightly. For instance, X_j for $j \geq 2$ changes from $X_2 = \pi\Pi_2^{T1} - c_2$ as given in (8) to

$$X_2 = \pi\Pi_2^{T1} - c_2 + \pi\delta(\Pi_1^{T0} - \Pi_2^{T0}) \quad \text{if } \kappa \geq \kappa^* \tag{10}$$

but are the same for $\kappa < \kappa^*$ (and similar changes in X_1 for the two subcases when $\theta \leq \theta^{**}$). Thus, the second (or j-th challenger) must be paid an additional amount equal to the discounted expected

value of incumbency, $\pi\delta(\Pi_1^{T0} - \Pi_2^{T0})$ with similar small increase in the X_1 payment for relevant subcase. Compared to the no incumbency advantage case, this change increases the parameter space over which P4D deals are not possible (area marked as 'IV - No Deal' increases) since the threshold moves to the right. But more importantly, parameter space over which P4D deals with one challenger (region 'III- P2D Pay Only First') does not shrink as J increases as was the case when there was no incumbency advantage (and as in the previous case, region II shrinks and IV increases with J). Thus P4D deals are still possible under the FSC system, though less so if there is also an incumbency advantage.

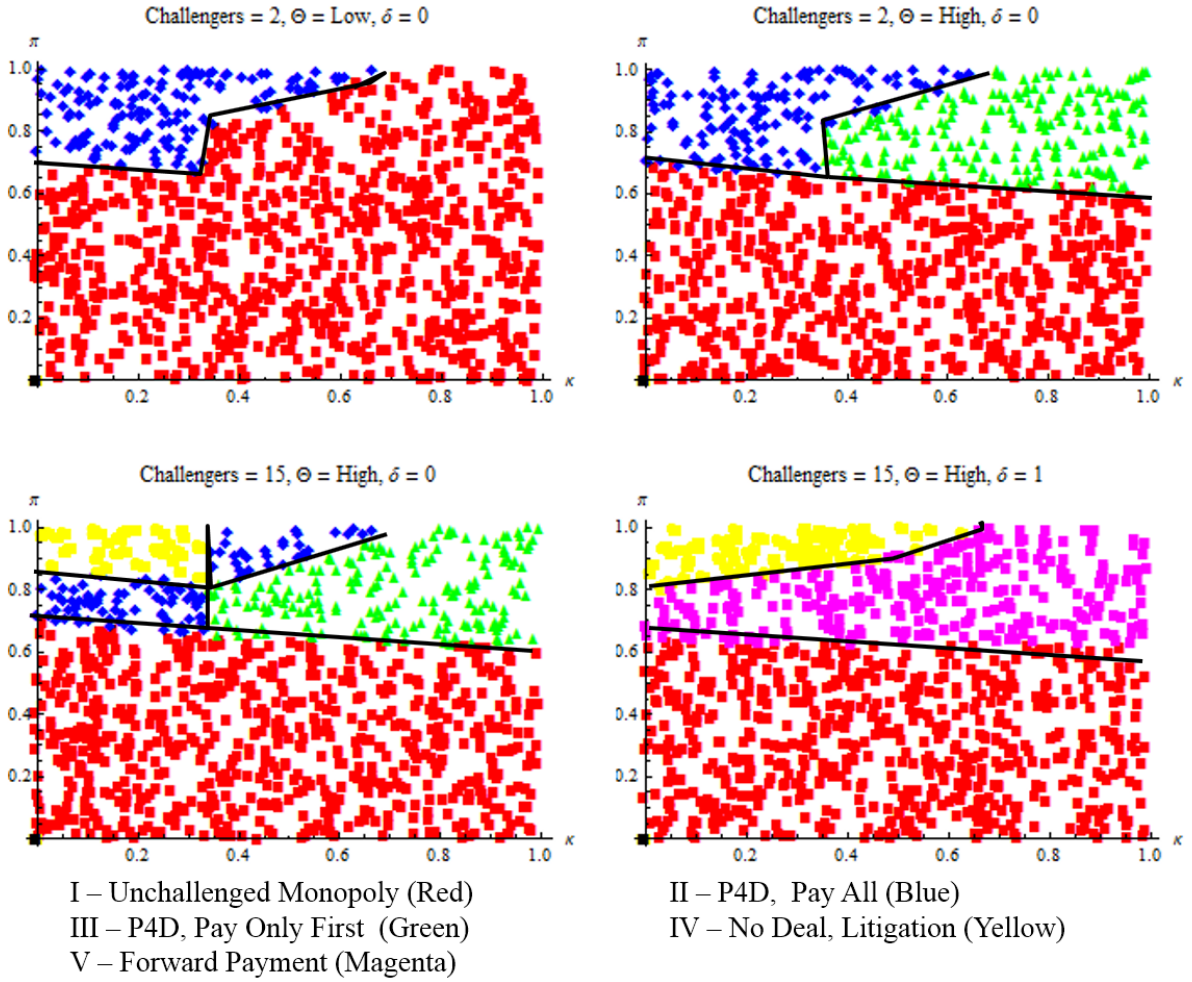


FIGURE A-11. Agreement outcomes with Risk Averse Challengers

A.3.6. *Risk Aversion.* Generic firms may be (more) risk averse than branded firms and hence susceptible to settling with the originator than taking a chance in court. In fact, they may not even mount a challenge unless the patent is sufficiently weak. To check this, we modified the payoffs in the game tree to be exponential utility function r of net profits $\widetilde{V}_j^{T\#}$ for generic firms. Specifically, we use the form

$$r(\widetilde{V}_j^{T\#}) = \begin{cases} (1 - e^{-a\widetilde{V}_j^{T\#}})/a & \text{if } a \neq 0 \\ \widetilde{V}_j^{T\#} & \text{if } a = 0 \end{cases} \tag{11}$$

where a is a parameter for risk aversion ($a > 0$ for risk aversion and $a = 0$ for risk neutrality). Exponential utility implies constant absolute risk aversion (CARA), with coefficient of absolute risk aversion equal to the constant a above.

The results shown in Figure A-11 are for the cases analyzed in the main paper in Section (4.4) and mimic results shown in Figure 2, but computed when $a = .005$ for the generic firms. Even with such a slight amount of curvature, we find that (1) payments to challengers reduce, (2) it is easier to reach P4D deals as the area for 'No Deal, Litigation' shrinks, (3) generics are less likely to bring challenges in the first place as area marked as 'Unchallenged Monopoly' increases.

A.3.7. *Welfare.* For the baseline case as well as all the policy options considered above, we can compute consumer surplus (CS) in the first period ($\delta = 0$) for each agreement outcome where the market structure is either a monopoly (no challenge or a P4D deal), or a duopoly with probability π if litigation results in patent invalidation (if the threat is not credible and litigation results in a win for the challenger, the branded firm does not launch an AG).[23] Thus, following Vives (1984), we can compute CS using the utility function of a representative consumer net of total expenditures on the drugs, i.e., $U(\mathbf{q}) - \mathbf{p} \cdot \mathbf{q} = \boldsymbol{\alpha}\mathbf{q} - \frac{1}{2}\mathbf{q}'\boldsymbol{\Sigma}\mathbf{q} - \mathbf{p} \cdot \mathbf{q}$ and computed at equilibrium quantities and prices for the given agreement outcome in the κ-π space.[24] Figure A-12 compares the level of CS in the baseline case (agreement outcomes with 15 challengers and $\delta = 0$, corresponding to panel 3 in Figure 2) with CS under the policy option considered above (no AG against a winning challenger).

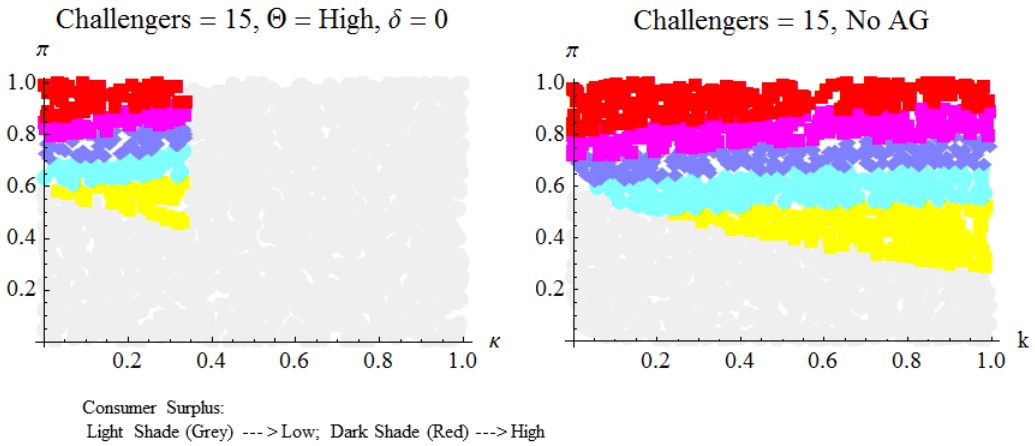


FIGURE A-12. Consumer Surplus

Consumer surplus is lowest when there is a monopoly (grey area) and equals weighted average of CS from monopoly and duopoly when there is litigation. In this region, expected CS increases with π since weaker patents are more likely to result in a duopoly. Under the considered policy option of No AG against a winning challenger, litigation becomes possible in the north east region of κ-π area, and consequently expected consumer welfare in this region increases. By comparison, for low values of π there is no change in CS from the adoption of this policy as strong patents continue to go unchallenged.

[23]Note that because of the constant market size restriction (i.e. that $a_0^{(T)} + a_1^{(T)} + a_2^{(T)} = a_0^{(D)} + a_1^{(D)} = a_0^{(M)}$) we can not directly compare consumer surplus across market structures using the utility function used here. However, consumer welfare calculations within a market structure can be derived as a function of κ, and compared to the monopoly case where WTP is constant.

[24]For instance, if $\kappa = 0$, then if litigation results in a a duopoly, our consumer surplus is given by $\frac{\beta^2}{(2\beta-\gamma)^2(\beta+\gamma)}\alpha_0^2$, which is precisely the expression for consumer surplus in Vives (1984) for Nash-Bertand duopoly case.

APPENDIX B. (SUPPLEMENTARY MATERIAL)

B.1. **Institutional Details.** The Food and Drug Administration (FDA) in the US and the European Medicines Agency (EMA) in the EU (or national medicinal agencies) are responsible for granting market authorization (MA) for drugs.[25] When applying to the EMA a firm can choose to apply via the community authorization procedure (CAP), where a single application can be used for authorization in multiple jurisdictions, or it can choose to obtain market authorization from a national agency directly and obtain authorization for that member state only. Alternatively, if the drug is already approved in one member state, the firm can apply for the mutual recognition procedure (MRP) to gain marketing approval in other member states. Finally, if no national market authorization exists, the firm can also use the decentralized procedure (DCP), which allows for submission of the application in select multiple member states, and where one country is designated as a reference member state. All in all, there are three different procedures (CAP, MRP and DCP) for gaining marketing authorization with the EMA or via 27 national medical agencies.

In the first instance, original drugs are protected from direct competition from generics via patents, which are granted for 20 years and confer monopoly rights to the originators. In the US, the originator lists the relevant patents with the FDA when filing for a New Drug Application (NDA), while in the EU a similar 'full application' is filed with the EMA but without any patent linkage. The drug approval process for new drugs lasts several years, involving multiple phases of clinical trials establishing safety and efficacy. However, since these cut the effective exclusive market life of the patented drug significantly, both the US and EU provide non-patent exclusivity to the originator to compensate for these delays. In the US, a market exclusivity period for the originator was introduced as part of the Hatch-Waxman Act, where the originator is protected from generic competition via the 'data exclusivity' period – a period during which a generic firm cannot rely on the original drug's safety and efficacy to file its own application. As per the provisions of the Act, a generic can forego clinical trials, citing safety and efficacy already established by the originator's reference drug, and file instead for bio-equivalence under the abbreviated new drugs application (ANDA) procedure, but not during the data exclusivity period. Testing and establishing bio-equivalence is also expensive and time consuming, but not as much as the clinical trials required when filing an NDA application (see Appelt (2015) for some recent estimates).

The Act also allows the generic firm to use the patented drug for testing bio-equivalence and developing an ANDA application without infringing the patent so that the ANDA application can be filed on the day the data exclusivity expires (this is the so-called Bolar exception). The data exclusivity period is five years for drugs classified as New Molecular Entity (NME), three years for new formulations (which also carry a patent but not on the molecule), and seven years for orphan drugs. Six month extensions can be added on as pediatric exclusivity when the firm conducted and submitted pediatric studies in response to request from the FDA. The five year exclusivity is cut to four years if the generic files under paragraph IV citing that the patent is either not valid, or will not be infringed.[26] In the latter case, the FDA informs the originator, and if the originator objects on grounds of patent infringement within 45 days, a one time 30-month stay order for generic entry comes into effect to allow the courts time to resolve patent litigation. Thus, generic entry typically takes place after resolution of patent litigation or settlement. Finally, as mentioned earlier, the first

[25]Information regarding market authorization rules, patent litigation, and other regulation in the pharmaceutical industry is well documented and hence we don't provide individual citations. Readers interested in further details on EU/US regulations and differences there in, are referred to (among others) Graham et al. (2002), Harhoff (2009), Glowicka et al. (2009), Hancher (2010) and Gürkaynak et al. (2014).

[26]For NME exclusivity, an ANDA application cannot be filed for first five years (or four years if it is para IV challenge), but for other cases, it can be filed but not approved by the FDA until the exclusivity period runs out.

filer is entitled to a 180-day market exclusivity period against other generics and is entitled to it even if it settles with the branded firm (also, the exclusivity does not pass over to a later ANDA filer if the first filer settles or loses the case). Further, as documented by Hemphill (2009), the first filer may not be just a single generic firm, as all firms that file on the same first day are awarded the 180-day exclusivity against other generics. Multiple filings on the same day can happen due to the Bolar exception since generics can start preparing for the ANDA filing during the exclusivity period.

In the EU there are two routes available to the originators to extend the exclusive marketing of their products from generic competition. The first, available since 1992, is the Supplementary Protection Certificate (SPC) available for medicinal products, which allows originators to extend the original patent for up to five years after the expiration of the original patent, or fifteen years from the first marketing authorization in the EU, whichever is less. While all member states provide SPC, there is no cross-border recognition, and hence the application has to be filed in each country where the originator wants to enforce and extend the patent life (Hancher, 2010, Graham et al., 2002). Further, patent infringement and validity fall under the jurisdiction of national courts, and hence patent-holders (or parties seeking to revoke granted patents) may have to enter into litigation in multiple countries resulting in duplication of cases (Harhoff, 2009, EC, 2009).[27]

Second, like the US, there is a data exclusivity period which was also introduced in 1984 as part of the mutual recognition procedure for drug approval in the EU (prior to that, drug approval was at the national level and with varying rules), and similar to ANDA, generics can file an 'abridged' application. Initially, data exclusivity extended either to six years from initial market authorization date, or ten years, depending on the member state, and did not include the Bolar provision to allow for use of patented drug for clinical studies.[28] Further, some member states opted not to allow for data exclusivity to extend beyond the patent expiration of the original product. In 2005, a new '8+2(+1)' exclusivity period was introduced which, (i) added the Bolar provision, and (ii) provided unified rules of exclusivity across member states – eight years of data exclusivity during which a generic cannot file for an abridged application, plus two additional years of market exclusivity, i.e., the generic may file the abridged application but not market the drug, and a final one additional year of market exclusivity for new indication(s) if they constitute a significant clinical benefit.

Thus both the US and EU provide data exclusivity periods during which the generic drugs cannot enter the market but entry can take place afterwards as long as there are no patents protecting the drug. In turn this implies that while there is no 30-month automatic stay order with the EMA, the branded firm can obtain interim injunction from the national court(s) to prevent generic entry until the litigation case is resolved. In the *Pharmaceutical Sector Inquiry* (SI) by the DG Competition, the average patent litigation was 2.8 years, and interim injunctions were granted in 44% of the cases lasting on average 18 months (EC, 2009, pp.229-238). Even if there are no injunctions, the report also notes that some health authorities responsible for pricing and reimbursement of medical products can require certification from the drug manufacturers that no patent is infringed, and hence in the EU (as in the US) generic entry would mostly take place *after* patent litigation is resolved.

[27]However, patent 'opposition' can be filed at at EU level at the European Patent Office (EPO) but must be initiated within the first nine months from the grant of the patent.

[28]Austria, Denmark, Finland, Ireland, Portugal, Spain, Greece, Poland, the Czech Republic, Hungary, Lithuania, Latvia, Slovenia, Slovakia, Malta, Estonia, Cyprus, Norway, Liechtenstein and Iceland provided six years of data exclusivity while Belgium, Germany, France, Italy, the Netherlands, Sweden, the UK and Luxembourg had ten year of exclusivity for the originator.

Finally, while there is also no automatic 180-day exclusivity period for the first generic entrant in the EU, delays in the drug approval process at the EMA or by national authorization agencies may provide the first generic entrant a short lived duopoly period. As pointed out in the SI report, this can happen when a national authorization agency has to act as a reference member state in MRP/DCP application with the EMA, and the work load at the national agency is high enough for it not to be able to process additional applications for another one or two years. The report notes that in 2008-09, several national agencies were already 'fully booked', that according to some generic companies they had to 'book 18 months in advance to get a slot for a product', and if they experienced any delay in development, they had to miss a whole year (EC, 2009, p.465). Some generic firms also reported that these bottlenecks were due to 'misuse of procedures by some applicants, who make "unnecessary" or parallel bookings, possibly also to delay access for other applicants'. While these delays apply to all firms and not just the second generic challengers, they can, nonetheless, create a wedge between the entry dates of the first and later generic challengers so as to create short duopoly periods for the first generic entrant.

B.2. **Differentiated Products and First Mover Advantage.** In this appendix we model demand with differentiated products and parameterize the first mover advantage (FMA) for the first generic via the demand curves and derive equilibrium profits.

B.2.1. *Market Demand Curves.* Following Singh and Vives (1984), we use a quadratic (strictly concave) utility function for a representative consumer to derive linear demand functions for differentiated products, but where differentiation exists up to the third product (second generic product), i.e., products $2, \ldots, J+1$ are homogenous with respect to each other. Thus, let $U(\mathbf{q}) = \boldsymbol{\alpha}\mathbf{q} - \frac{1}{2}\mathbf{q}'\boldsymbol{\Sigma}\mathbf{q}$ where the vector $\boldsymbol{\alpha}$ specifies the maximum willingness-to-pay (WTP) for the brand, generic 1, generic 2, and so on. In a triopoly $\boldsymbol{\alpha} = (\alpha_0^{(T)}, \alpha_1^{(T)}, \alpha_2^{(T)})$, while in a monopoly $\boldsymbol{\alpha} = \alpha_0^{(M)}$ (the branded firm), and similarly $\boldsymbol{\alpha} = (\alpha_0^{(D)}, \alpha_1^{(D)})$ in a duopoly between the branded and the generic entrant. When there are more than three firms in the market, we make the simplifying assumption that the market structure is approximated by a triopoly where the second generic is a collective sum of all the remaining identical generic firms, and thus $\alpha_2^{(T)} = \sum_{j=2}^{J} \alpha_j^{(N)}$. Similarly, $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix and we parameterize it with just two terms, β on the leading diagonal, and γ as the term on off-diagonals so that, in a triopoly, $\boldsymbol{\Sigma} = \begin{bmatrix} \beta & \gamma & \gamma \\ \gamma & \beta & \gamma \\ \gamma & \gamma & \beta \end{bmatrix}$ where $\beta > 0$. As such γ can be negative, positive or zero corresponding to complementary, substitute or un-related products but we focus on the case when the drugs are substitutes, and hence set $\gamma > 0$. In the case of a duopoly, $\boldsymbol{\Sigma}$ is a two by two matrix with similar terms, while in the case of a monopoly, it is a scalar equal to β. While $\boldsymbol{\Sigma}$ appears very restrictive with just two parameters, it suffices for our purpose, as we wish to highlight the role of the first mover advantage for the first generic in determining the outcomes in the earlier game, which we capture via the WTP parameters $\alpha_1^{(T)}$ and $\alpha_2^{(T)}$ in relation to $\alpha_0^{(T)}$ for the branded firm. Our motivation for this choice of modeling comes from the fact that patients (and physicians and pharmacists) may view the branded drug to be of a different quality than the generic, but without a price differential they may be less willing to switch from the first to the second generic, i.e., inherently view the latter generic(s) to be of lower quality (Hollis, 2002). An alternative would be to model FMA by changing either the parameters that directly affect the demand sensitivity of own price (so that the leading diagonals are not all equal to β but instead given by β_j) or by not making all the off-diagonals equal, particularly $\gamma_{01} = \gamma_{10} \neq \gamma_{02} = \gamma_{20}$. However, these latter parameters are better suited to capture the degree of product differentiation via price effects, and hence we keep this matrix simple, and simply note that the price elasticities

will be defined by both sets of parameters (i.e., $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$) and hence the cross-price effects need not be symmetric.

To derive demand functions that correspond to a utility maximization problem, it must be true that $\boldsymbol{\Sigma}$ is positive definite, which in turn requires that

$$\beta - \gamma > 0 \qquad \text{and} \qquad \beta + 2\gamma > 0 \tag{12}$$

where the restrictions arise because $|\boldsymbol{\Sigma}| = (\beta-\gamma)^2(\beta+2\gamma)$ and the eigenvalues are $\{\beta-\gamma, \beta-\gamma, \beta+2\gamma\}$. The inverse and direct demand functions are then given by $\mathbf{P}(\mathbf{q}) = \boldsymbol{\alpha}' - \boldsymbol{\Sigma}\mathbf{q}$ and $\mathbf{D}(\mathbf{p}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha}' - \mathbf{p})$. Solving explicitly, the inverse and direct demand functions for the triopoly are,

$$\begin{aligned} p_0 &= \alpha_0^{(T)} - \beta q_0 - \gamma q_1 - \gamma q_2, & q_0 &= a_0^{(T)} - b p_0 + c p_1 + c p_2 \\ p_1 &= \alpha_1^{(T)} - \gamma q_0 - \beta q_1 - \gamma q_2, & q_1 &= a_1^{(T)} + c p_0 - b p_1 + c p_2 \\ p_2 &= \alpha_2^{(T)} - \gamma q_0 - \gamma q_1 - \beta q_2, & q_2 &= a_2^{(T)} + c p_0 + c p_1 - b p_2. \end{aligned} \tag{13}$$

In the equation above, the parameters a, b, c represent the relative size of the market and price coefficients and are related to the primitives of the model by

$$\begin{aligned} a_0^{(T)} &= \left[\alpha_0^{(T)}(\beta+\gamma) - \gamma(\alpha_1^{(T)} + \alpha_2^{(T)})\right]/d \\ a_1^{(T)} &= \left[\alpha_1^{(T)}(\beta+\gamma) - \gamma(\alpha_0^{(T)} + \alpha_2^{(T)})\right]/d \\ a_2^{(T)} &= \left[\alpha_2^{(T)}(\beta+\gamma) - \gamma(\alpha_0^{(T)} + \alpha_1^{(T)})\right]/d \\ \text{where } b &= (\beta+\gamma)/d, \qquad c = \gamma/d, \qquad \text{and } d = (\beta-\gamma)(\beta+2\gamma). \end{aligned} \tag{14}$$

Since d is positive (see restriction (12)), it also implies that $b > 0$ and $c > 0$. Note that if we allowed complementarities in the model so that $\gamma < 0$ and hence $c < 0$, we would then explicitly require $\beta + \gamma > 0$ for downward sloping demand curves. The demand equations in the case of duopoly and monopoly are similar to the linear structure above but omitted in the interest of space. An additional condition under duopoly is that (12) is modified to $\beta - \gamma > 0$ and $\beta + \gamma > 0$ rather than $\beta + 2\gamma > 0$ under triopoly (but these are automatically satisfied in a duopoly if they are already satisfied in a triopoly).

B.2.2. *Willingness to Pay.* To ensure positive demand curves, the intercepts $a_j^{(T)}$ must be positive (equivalently, we can impose second order conditions for profit maximizing which would impose similar restrictions on demand parameters). Positive demand implies that WTP for the two generics $\{\alpha_1^{(T)}, \alpha_2^{(T)}\}$ be such that

$$\begin{aligned} &\alpha_2^{(T)} < \left(\frac{\beta+\gamma}{\gamma}\right)\alpha_0^{(T)} - \alpha_1^{(T)}, \qquad \alpha_2^{(T)} < \left(\frac{\beta+\gamma}{\gamma}\right)\alpha_1^{(T)} - \alpha_0^{(T)} \text{ and,} \\ &\alpha_2^{(T)} > \left(\frac{\gamma}{\beta+\gamma}\right)\alpha_0^{(T)} + \left(\frac{\gamma}{\beta+\gamma}\right)\alpha_1^{(T)}. \end{aligned} \tag{15}$$

The shaded region in Figure B-13 shows the allowed range for WTP parameters for the two generics given the WTP for the branded drug $\alpha_0^{(T)}$ (outside the range the problem is not of any economic interest). The 45° line (given by $\alpha_1^{(T)} = \alpha_2^{(T)}$ but within the region), indicates that a patient's willingness to pay for the two drugs is equal, but increasing relative to the branded drug as we move further away from the origin. This in turn implies that the potential market size for the generics is equal (i.e., $a_1^{(T)} = a_2^{(T)}$, see (14)) on the line, but increases in magnitude as we move further away from the origin. All points off the 45° line increase the WTP for one or the other generic (and consequently imply a larger potential market for that generic). We choose movements

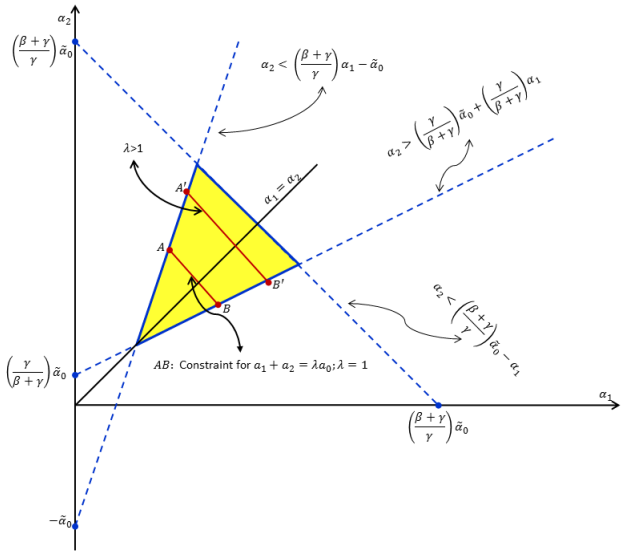


FIGURE B-13. WTP for Generics in a Triopoly

along line segments such as $A'B'$ to parameterize first mover advantage, where all points on the line segment fix total potential market size of generics as a proportion of the branded market. Specifically, along all points of $A'B'$, we have $a_1^{(T)} + a_2^{(T)} = \lambda a_0^{(T)}$, where $\lambda > 0$. Then in terms of WTP of the branded drug, points on $A'B'$ are parameterized as

$$\begin{aligned} \alpha_1^{(T)} &= (1-\kappa)\left[\frac{(2+\lambda)\gamma + \beta\lambda}{2(\beta+\lambda\gamma)}\right]\alpha_0^{(T)} + \kappa\left[\frac{\gamma+\beta\lambda}{\beta+\lambda\gamma}\right]\alpha_0^{(T)}, \\ \alpha_2^{(T)} &= (1-\kappa)\left[\frac{(2+\lambda)\gamma + \beta\lambda}{2(\beta+\lambda\gamma)}\right]\alpha_0^{(T)} + \kappa\left[\frac{(1+\lambda)\gamma}{\beta+\lambda\gamma}\right]\alpha_0^{(T)}, \end{aligned} \tag{16}$$

for $\kappa \in [0,1]$, where $\kappa = 0$ implies neither generic has a first mover advantage (corresponds to the point on the 45° line) and $\kappa = 1$ means that the first generic has the maximum first mover advantage (allowing negative values of κ up to negative one allows for modeling second mover advantage but is not of interest to us). On the other hand, the λ parameter sets the relative market size between the generic and branded segments of the market, and is determined by the WTP for generics relative to that of the branded product.

To compare outcomes (prices, quantities, and profits) across market structures (triopoly, monopoly or duopoly), we impose the restriction that the total (potential) market size under the three structures is the same. Thus, we assume that the introduction of generics to the market does not increase the potential set of patients *per se*, meaning no new patients exist that can use the drug, though in equilibrium the actual number of patients that consume the drug may increase due to lower prices if existing patients were originally priced out and hence, for comparison, we impose

$$a_0^{(T)} + a_1^{(T)} + a_2^{(T)} = (1+\lambda)a_0^{(T)} = a_0^{(D)} + a_1^{(D)} = a_0^{(M)}. \tag{17}$$

In turn, this implies that if the WTP for the branded drug in a monopoly is normalized to $\alpha_0^{(M)} = \widetilde{\alpha}_0$, then in a triopoly,

$$\alpha_0^{(T)} = \frac{\beta+\gamma\lambda}{\beta(1+\lambda)}\widetilde{\alpha}_0 \tag{18}$$

i.e., the willingness-to-pay for the branded drug would be lower in a triopoly.[29] In a duopoly, we allow the solo generic to have the same WTP as the first generic entrant in a triopoly, while making sure that the potential market size is constant. Specifically, let $\alpha_1^{(D)} = \alpha_1^{(T)}$ and set $\alpha_0^{(D)}$ such that (17) holds, which gives

$$\alpha_0^{(D)} = \frac{\beta + \gamma}{\beta}\widetilde{\alpha}_0 - \alpha_1^{(D)} \qquad \text{and} \qquad \alpha_1^{(D)} = \alpha_1^{(T)}. \tag{19}$$

An implication of the restriction that the potential marketsize is constant (i.e. $a_0^{(T)} + a_1^{(T)} + a_2^{(T)} = a_0^{(D)} + a_1^{(D)} = a_0^{(M)}$), is that we can not directly compare consumer surplus across market structures using the above utility function. This is because the constant market size restriction implies that $\boldsymbol{\alpha}$ is different in duopoly vs triopoly and hence $\alpha_0^T \neq \alpha_0^D \neq \alpha_0^M$, i.e., maximum willingness to pay for a branded product itself is changing (decreasing) when there are more generic products in the market. Nonetheless, welfare calculations within a market structure can be derived as a function of κ, and compared to the monopoly case where WTP is constant.

B.2.3. *Price Competition.* We model competition as Nash-Bertand with differentiated products. Consider first the case where all three firms engage in price competition in a competitive triopoly where there are no authorized generics, i.e. configuration T0. Then the profit maximizing equilibrium prices are determined by

$$\begin{aligned} p &= c + \Omega^{-1} D(p_0, p_1, p_2) \text{ and where } \Omega \text{ is a } 3 \times 3 \text{ matrix such that} \\ \Omega_{ij} &= -O_{ij} \frac{\partial D_j(\cdot)}{\partial p_i}. \end{aligned} \tag{20}$$

In the equation above, O_{ij} are terms of the 'ownership' matrix, set equal to the identity matrix for the baseline case of a competitive triopoly (Nevo, 1998). Triopoly outcomes in other cases (authorized generics) are computed similarly but by adjusting the terms of the ownership matrix. For instance, when the branded firm launches an AG via the first challenger and competes with the second challenger (T1), equilibrium prices are computed by setting the off-diagonal terms for the branded and the first generic equal to one in the ownership matrix to allow for joint profit maximization between these two firms.[30] In a duopoly, the pricing equation is similar except that dimensionality is reduced by one, and the ownership matrix is either equal to an identity matrix (in the D0 competitive duopoly case) or all terms are equal to one (in the D1 duopoly where the branded firm has launched an AG). Computation of equilibrium prices allows computation of quantities and firm profits.

[29] This follows from the inverse demand function in monopoly defined equivalently as $p_0 = \alpha_0^{(M)} - \beta q_0$, which gives the demand function as $q_0 = a_0^{(M)} - b_0^{(M)} p_0$ where $a_0^{(M)} = \alpha_0^{(M)}/\beta$ and $b_0^{(M)} = 1/\beta$ and then using substitution and simplification from earlier relations. Note that as long as $\beta > \gamma$, the WTP of the branded in triopoly is always lower than that in monopoly for all $\lambda > 0$. Further, it is decreasing function of λ.

[30] Similarly, our model allows for a fully collusive triopoly, i.e., the branded firm launches two AGs, and is in a 'T2', all terms of the ownership matrix are set to one.